

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 06:44:34 ; Search time 4634.01 Seconds

(without alignments)
17019.508 Million cell updates/sec

Title: US-10-026-994-1

Perfect score: 2710
Sequence: 1 ccagcgctccgagcgatgctc.....ctatgaaaaaaaaaaaaa 2710

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb.ba:*
- 2: gb.htg:*
- 3: gb.in:*
- 4: gb.om:*
- 5: gb.ov:*
- 6: gb.pat:*
- 7: gb.ph:*
- 8: gb.pl:*
- 9: gb.pr:*
- 10: gb.ro:*
- 11: gb.sts:*
- 12: gb.sy:*
- 13: gb.un:*
- 14: gb.vl:*
- 15: em.ba:*
- 16: em.fun:*
- 17: em.hum:*
- 18: em.in:*
- 19: em.mu:*
- 20: em.om:*
- 21: em.or:*
- 22: em.ov:*
- 23: em.pat:*
- 24: em.ph:*
- 25: em.pl:*
- 26: em.ro:*
- 27: em.sts:*
- 28: em.un:*
- 29: em.vl:*
- 30: em.htg.hum:*
- 31: em.htg.inv:*
- 32: em.htg.other:*
- 33: em.htg.mus:*
- 34: em.htg.pln:*
- 35: em.htg.rod:*
- 36: em.htg.mam:*
- 37: em.htg.vrt:*
- 38: em.sy:*
- 39: em.htgo.hum:*
- 40: em.htgo.mus:*
- 41: em.htgo.other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	623.8	23.0	2745	8	AB015511 Aspergill
2	481.4	17.8	2579	8	AJ292929 Agaricus
3	436.2	16.1	41806	1	SCSC7 AL031515 Streptomy
4	334.6	12.3	3359	8	AY040839 Aspergill
5	243	9.0	3262	1	AF078038 Caldcell
6	203.6	7.5	12732	1	AE007608 Eimeri
7	127.4	4.7	3030	8	VC4429230
8	116.8	4.3	11910	1	AE012276
9	105.4	3.9	14520	1	AE011809
10	98.6	3.6	58930	2	AC098321
11	93	3.4	88203	5	AC097628
12	92.8	3.4	232869	10	AL603925
13	92.2	3.4	16150	1	MTV044
14	92.2	3.4	16168	1	AE006985
15	92	3.4	181179	9	AL161725
16	91.2	3.4	13511	1	AE007164
17	89.4	3.3	87076	9	AC005918
18	88	3.2	15311	1	AE007164
19	87.8	3.2	126599	2	AP003816
20	87.8	3.2	167254	9	CNS057DS
21	87.2	3.2	65140	6	AX211705
22	87.2	3.2	123580	1	AF263912
23	87.2	3.2	125401	6	AX211739
24	86.2	3.2	200350	2	AC095469
25	86	3.2	169585	2	AC078821
26	85.8	3.2	14860	1	AE007093
27	85.8	3.2	63033	1	MTV008
28	85.8	3.2	154292	8	AP003437
29	85.8	3.2	15574	2	AC091090
30	85.6	3.2	47852	1	MTV023
31	85.2	3.1	347750	1	AP002998
32	85	3.1	15348	1	AE007163
33	84.8	3.1	42655	1	SC7H2
34	84.8	3.0	120185	1	MTV023
35	81.2	3.0	198220	8	AC007138
36	81.2	3.0	157922	2	ATCHRIV5
37	80	3.0	157922	2	AC123764
38	79.8	2.9	20701	1	AE006958
39	79.8	2.9	29550	1	MTV039
40	79.8	2.9	40740	9	AC027349
41	77.6	2.9	167390	9	AC007263
42	77.4	2.8	38425	1	SCF55
43	76.8	2.8	1833	6	A37836
44	76.8	2.8	1833	6	AR069874
45	76.8	2.8	1833	6	AR099270

ALIGNMENTS

RESULT 1
LOCUS AB015511 2745 bp mRNA linear PLN 20-JUN-1998
DEFINITION Aspergillus aculeatus mRNA for Avicelase III, complete cds.
ACCESSION AB015511
VERSION AB015511.1 GI:3242654
KEYWORDS Avicelase III.
SOURCE Aspergillus aculeatus cDNA to mRNA.
ORGANISM Aspergillus aculeatus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 2745)
AUTHORS Arai, M., Takada, G., Kawaguchi, T. and Sumitani, J.
TITLE Avicelase III from Aspergillus aculeatus
JOURNAL Published Only in Database (1998)


```

OY 467 GCCGAGAGCGTCTGGCTGATCGAGCCCACTCCAACTCATCTTGGTGTGCG 526
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 472 ATGGAGAGCGCTCTGCTGATCCCAATCTGACAGCATCTCTTATTTATGGCGCTCG 521
OY 527 TCAGGAAGCGCTCTGGAAGTCTACGAGCGCGGCTGACTTTTCCAGAGCTCTCG 586
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 532 AGTGTAAACGCTCTGGAAGTCAACAAATTCGGGTGAGCTGAGCAAGTAACAGC 591
OY 587 TTCACGCACTGGAGCTGATCCAGACCGAGTATTCAGAGCTCAACAGAGAC 646
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 592 TTACTGACACTGGGTCTTCTGTTCCGATCTCATAGACACGGCTGAAATTCGAT 651
OY 647 AACCAAGACTATGGGTATGCTTGCATCAACAGCAAGACGAGCGGGAGCCAGC 706
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 652 AAGATGGGATCCGATGGGTATACGATAAAGCGTACGCTCTGTTCCGCTACT 711
OY 707 TCTGATCTTTTGTGGACGCGTGAATACATACCTTCACTTATCTGATGACGAT 766
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 712 CCCCCTATATTTTGTGGGCTGGAACAGGCTCAAAAGCATTTATATCTAGCAAT 771
OY 767 GCCGCTCCAGTGTGAGTGTACCGGGGACAGGAAATCTTCTCAAGGCG 826
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 772 GGGGTTTCTCTTGGAGTGGCGGTGCAAGTCAACAACTGCTTCCGCACTAAGGC 831
OY 827 AAACAGCAGCCAGAGAAAGCCTGTATCTGACCTATTCCGATGGACAGGCGCTAT 886
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 832 GTTCTATCCCTTCTGTGAGGCACTTATATCACTTACCTGATGGTGGCTCTTAC 891
OY 887 GATGCACTTGGCTGAGTGTGAGGTAGCATGACATGGGGAACCTTGGAAAGCATC 946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 892 GATGCACTGAGTGTGAGGTAGCATGACATGGGGAACCTTGGAAAGCATC 951
OY 947 ACCCTGTCTGTGATCATATCTTGTGCTTGTGAGGCGCTTGGCTTGTGCAA 1006
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 952 ACACCGTCAAGGAGCATCTCTTGTGTTGCGAGCTGACGCTTATGATACAG 1011
OY 1007 AACCCAGGAACCTTGTGTGCTTCTTGAATCTTGGTGGCAATGCTGACGCTT 1066
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1012 AAAAAGGCGCTCATGTTGAGCCCTCAACTCGTGGCTTGTGATGATATTC 1071
OY 1067 CGGTGCACTCTGTGAGCAACATGAGCCGATCTGGGCGAGGAGCTATTCGCACT 1126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1072 CGGTGCACTCTGTGAGCAACATGAGCCGATCTGGGCGAGGAGCTATTCGCACT 1131
OY 1127 GAGACCTATTACTAGACATCTCACTCCCAAGACCGCTGATCAAGAACTTATC 1186
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1132 CTCAACAAATCTAGATTAACGCTGCGCTGGCGCTGATAGACCAATCTGTT 1191
OY 1187 GATGTGAGAGGAGTACCGCTCGATGCTCATCAAGGCGCTGGCTGATGATGAG 1246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1192 GATGTGAGAGGAGTACCGCTCGATGCTCATCAAGGCGCTGGCTGATGATGAG 1253
OY 1247 TCTGTGAGATGAGCAACACGACAGCAACCATGCTTACGCAACGGAATGACATC 1306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1234 TCATCTATCATCTGATCTTCTGACATGCAATCATTTGGCTTACGCAATGAAATC 1293
OY 1307 TTGGGCGGCGACATCTCACTAGGAGACCGGCAACATGTGCAATCACTACTG 1366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1294 TATGGAAGTGTGATCTCTCAATATGAGACAGTCAACACATCAAGTGTGCTC 1353
OY 1367 GCAGAGGAGTGTGAGAAATCTCTGTCAGAGCTGCGCTGACCCGCGGAGAGAG 1426
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1354 GCAGATGCTGTGAGAAACTCTGTCAGAGCCCTTATTTCACTCTCTTGGAGCCCT 1413
OY 1427 CTATTTGGCGCACTCGAGAGCAACAGGCTTACCTTTCCAGAGAAAGCACTCGG 1486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1414 CTGTGTTTCTCTATTTGTGTGAGCTTTGGGGT-----TTGTCAACATCACTTAACCGGTG 1468
OY 1487 ACATGCGCGCAGACGCTGAGGCAACGCCCATATGGGCACTCTGAGAGCGTCTGACTAC 1546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1469 CACCCAGCGCCCA-----GTTTACCAATCCCAAGTGTGCTGCTGACATGATTC 1524
OY 1547 GCGGGAAGTCTGCAAGACGCTGCTCGGTGGCAACACCGCGGCAACG-----AA 1600

```

```

DB 1525 GCTGGCAATTAACCTAGTATGCTGTTCCATGTGATGATAGTACTAGTGCAAA 1584
OY 1601 CAGGTGGCCATCTCTGCCAGCGCGGCGGAGCTGAGATGATGATACCGCGCCAGCAG 1660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1585 CAGGTGGCCATCTCTGCCAGCTGATGCTGTTACTTGAATCAGATTTTGGAGCTCTGAT 1644
OY 1661 TCCATAGCGCGGCGGAGCTGATGCTTATGCGCGGAGCGGACAGATCTCTGTCGACC 1720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1645 AATGTTCAAGCGCGGAAGTGTGCAATCTCGCTGATGCGGCAATCATCTCTGCGCTACA 1704
OY 1721 GCTGTCTCGCGCGTCAAGCGCTTCCAGTTCCAGGCACTT 1761
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1705 AATGTTAAGGCTGATGTTGCGCAAAACAGGCAACGCTT 1745

RESULT 3
SC5C7/c.
LOCUS
DEFINITION
Streptomyces coelicolor cosmid 5C7.
ACCESSION
AL031515.2
VERSION
GI:20520758
KEYWORDS
ATP/GTP-binding protein; lyase; narg; narg; narg; nitrate
reductase; oxidoreductase; RNA polymerase sigma factor; secreted
cellulase; uvra-like protein.
SOURCE
Streptomyces coelicolor A3(2).
ORGANISM
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycetaceae; Streptomyces.
REFERENCE
1 (bases 1 to 41906)
Redenbach,M., Kiese,H.M., Denapalte,D., Eichner,A., Cullum,J.,
Kinashl,H. and Hopwood,D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
JOURNAL
MEDLINE
97000351
PUBMED
8643436
REFERENCE
2 (bases 1 to 41906)
Seeger,K.J. and Harris,D.
Unpublished
JOURNAL
3 (bases 1 to 41906)
Parkhill,J., Barrell,B.G. and Randsdram,M.A.
Direct Submission
Submitted (04-SEP-1998) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
On May 9, 2002 this sequence version replaced gi:3559990.
COMMENT
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/projects/S.coelicolor/) CDS are
numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
of Bibb et al., Gene 30:157-66(1994) as implemented at
http://www.nh.go.jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or (att)) which is preceded by an upstream

```

ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring subinserts. Cosmid 5C7 lies between 1B6 and 4B5 on the AseI-A genomic restriction fragment.

FEATURES
source

```
misc_feature
RBS
RBS
RBS
gene
CDS

/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 5C7"
2..110
/note="nominal overlap with cosmid 5C7 from 1 to 109"
complement(143..147)
/note="possible RBS upstream of SC5C7.01c"
337..340
/note="possible RBS upstream of SC5C7.02"
346..2736
/gene="SC5C7.02"
/note="SC06517"
346..2736
```

misc_feature	472..1701 /gene="SC5C7_02" /note="pfam match to entry PF00005 ABC_tran, ABC transporters, score 42.40, E-value 1.7e-10"
misc_feature	493..516 /gene="SC5C7_02" /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
misc_feature	1894..2628 /gene="SC5C7_02" /note="pfam match to entry PF00005 ABC_tran, ABC transporters, score 71.20, E-value 2.2e-17"
misc_feature	1915..1938 /gene="SC5C7_02" /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
gene	2804..3286 /gene="SC5C7_03" /note="SC06518"
DSDS	2804..3286

misc_feature

gene

CDS

```

/gene="SC5C7.04"
/note="SC5C7.04, possible lyase, similar to e.g.
IGUL, HAREN lactoylglutathione lyase (EC 4.4.1.5) (135 aa),
fasta scores: opt: 124 z-score: 215.2 E(): 9.7e-05, 29.08
identity in 131 aa overlap"
/codon_start=1
/transl_table=11
/product="putative lyase"
/protein_id="CAA20616.1"
/db_xref="GI:3559994"
/db_xref="SPTREMBL:O86701"
/translition="MDISLCFIVDDHDLALAFYRDVLGLEVRNDVGEGMAYWVL
GSPAPDDVIVLEPPLADPNASADKQAMADLLKGLHRLGITLTDVDAITFEHIVRAA
GAEVLEQEDVDQYGVRCACAFRDPAGNMVRLKLRPKG"
3979. 3982
/note="possible RBS upstream of SC5C7.05"
3995. .4795
/gene="SC5C7.05"
/note="SC06520"
3995. .4795

```

CDS

Query Match 16.1%; Score 436.2; DB 1; Length 41906;
Best Local Similarity 52.4%; Pred. No. 2.4e-50;
Matches 1156; Conservative 0; Mismatches 998; Indels 51; Gaps 7;

OY 112 GAAGCTCAACCTGGGCGGCGGCGGCTGCTGCTCCGCGCATCTATCCATCCCAAGAC 171
DB 34344 GAAGAACCCCGGATCCAGCGCGCGGCTTCTCCGCGCATCTATCCAGCGACGA 34285
OY 172 AAAAGCGGTAGATATGACAGAACAGATATTGGCGGCTTACCGCTCAACCGCAGCA 231
DB 34284 GAAGGACTGGCTTACCGCGCGGCGGCGGCTTACCGCGCTTACCGCGGAGAGGTC 34225
OY 232 CTATGAGACCGCGCTGACGATGGGATGCTGATATCCGCGCTGACAACTGGGCGAT 291
DB 34224 CCACACTGGAGCGCGCTCTCCGACACGTCGCGTGGGAGACCTGGGCGGCGACGCGCT 34165
OY 292 CGAGCGCTTGGCGCTTATGTCGCGAGCGATCAAAAGGTGATGCGCGCATCCGATGTA 351
DB 34164 GGTGGCTCTGCTCCGACGCGCGCTGATCCGAGCGGCGTCTACGCGGCGGCTGCGACGTA 34105
OY 352 TAGGACAGCTGGGATCGAGTATGAGGCGCATATTCGCTGCTGACAGCGCGCGCAAC 411
DB 34104 CACCAACGATGGGAGCCGACCAACGCGGCGGCTGCTGCTGCGCGCGACCGGCGGAG 34045
OY 412 GTGGCTTACCAACTTGGCTTCAAGTCGCGGCTTACATGCGCAGACCGGAGCGCG 471
DB 34044 CTGGGAGAAAGCGGACCTGCTTCAAGTCGCGGCGCAACCTGCGCGCGCGCATGGG 33985
OY 472 AGAGCGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 531
DB 33984 TGAGCGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33925
OY 532 AAAGCGCTTGGAAAGTCTAGCGAGCGGCGGCTGCTGCTTCCAAAGTCTGCTGCTGCT 591
DB 33924 CCAAGGCTTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33865
OY 592 GCGACCTGGGAGCTATCTCCAGACCGGAGTATTCAGCGGCTTACAAACGCGCAAGCA 651
DB 33864 GAACCCCGGAGAACTACCGCGAGCGGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 33805
OY 652 AGGACTGATGGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 711
DB 33804 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33745
OY 712 TATCTTGTGGACGCGCTGATACACTGCTTCTATGCTATGAGCAAGATGCGCG 771
DB 33744 GACCTCTACGCTGGGCTGCTGCGGAGAAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 33685
OY 772 CTGACCTGAGTGTCTTACCGGCGGAGCGAAGGAAATCTTCTCTACAAAGCGGAACT 831
DB 33684 CCGGACCTGGGAGCGGCTGCTGCGGAGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 33625
OY 832 GCAGCGCAGAGAGAGCGCTTGTATGCTATTCGATGGGCGCAGAGCGGCGTATGATG 891
DB 33624 GAGCGCGGAGAGAGCGCTTGTATGCTATTCGATGGGCGCAGAGCGGCGGCGGCTATGAGCG 33565
OY 892 CACACTTGGCTCAAGTGTGAGATGACGATTCAGAGGGAATTTGAAAGACATCAACCC 951
DB 33564 CCGGAGCGGCGGCTGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 33505
OY 952 TGTCTCTGAGTCAAGTATATCTTGGCTTGGGCGGCTTGGCTGCTGCTTGGCAAAAGCC 1011
DB 33504 GCGCGCGGAGAGCGGAGCGGCTTCAAGGCTTCAAGGCGCTTCAAGGCGGAGCGGAG 33445
OY 1012 AGGAACTGCTGCTTCTTGTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1071
DB 33444 GCGGAGCGGCTATGCGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33385
OY 1072 GACGAGCTTGGGAGCAAGTGAAGCGGATCTGGGCGTGGGCGAGCTATCGAGCTGAGAC 1131
DB 33384 GAGGAGAGCGGCGGAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33325
OY 1132 CTATTTACTACAGCATCTCAACTGCCAAAGCAGCGTGGATCAAGCAACTTTTATGATGT 1191

DB 33324 GAACCGCTTACACCATGAGACGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33265
OY 1192 GAGGAGCTTACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1251
DB 33264 GCGGCGGAG-----CAGACCGGAGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33220
OY 1252 CGAGATTGACCAACCGAGACAGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1311
DB 33219 GAGAGTGAACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33160
OY 1312 GCGGCGAGATCTTCAACCACTG---GACAGCGGCGCAGATGTGCTGCTGCTGCTGCTGCT 1368
DB 33159 CAGGAGAACCTGACGAACTGGGAGCGAGCGGCGGCGGAGCTTGGCTGCTGCTGCTGCTGCT 33100
OY 1369 AGAGCGCTGAGGAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1428
DB 33099 GCGGCGGCTTGGAGAGAGAGCGGCGCTTCAAGACGCTGCTGCTGCTGCTGCTGCTGCT 33040
OY 1429 ATTGGCGCAGTGGAGACGAGCAAGAGGCTTACCTTGTGCGAGGAAAGACCTGCGGAC 1488
DB 33039 GCTTACGCGCTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32985
OY 1489 ATGCGCGACAGGCTTGGGCAAGCGGCGCAGATGGGCGCAGCTGCTGCTGCTGCTGCTGCT 1348
DB 32984 ---CGCTGATGATGATACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32929
OY 1549 CCGGAACTGCGTCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1608
DB 32928 GAGAGCAAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32869
OY 1609 CATCTGCTCGAGCGGCGGCGGCGGAGTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1668
DB 32868 GTTCTCACGAGCAACCGGCGGCGGCGGAGTGGTGGGCGGCGGAGCGGCGGCTGCTGCTGCT 32809
OY 1669 CCGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1728
DB 32808 CCGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32749
OY 1729 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1785
DB 32748 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32689
OY 1786 GCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1845
DB 32688 GCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32629
OY 1846 GACTTTTACGTCAGCAAGAGCAACCGGCGCAGAGCTTCA--GCGCGGCGGCGGCGGCGGCA 1904
DB 32628 GAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32569
OY 1905 GCGGAGGAGATTCGCGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1964
DB 32568 GCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32509
OY 1965 CGACCGAG-----TCGGCATTTTCCGCTTCAACAGACTTGGGCGAGCAG 2007
DB 32508 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 32449
OY 2008 CTTTGGCGAGTCTC---ACCGCGCTGACCAACACTTACAGATGCGCTGCGGCTGGG 2064
DB 32448 CTTTACAGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32389
OY 2065 CTGAGGCTGCAACTGGAAGCTGATGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 2124
DB 32388 GCGGAGCGCTGCTTACCAAGCGCTTCTTCAACAGCGGAGAGATGGGCGGCTGCGGCGGAGCT 32329
OY 2125 CCGCAAGTGAAGAGCGGCGCTTCTGAGCGAGATTCAGAGGCTTCCAGGCGCTTGGGCTC 2184
DB 32328 CCGCTTACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 32269
OY 2185 CATGACAGCAAGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2244

Db	Accession	Definition	Version	Keywords	Organism	Source	Features
0Y	2245	CACGGCCGGGCGCTCTTTACGCTCAGGAACCGTCGGCGCG	2289				
Db	32208	CACGGCCGGCGCTGATCTACGGCACCTCGGACACGGCG	32164				
RESULT 4							
LOCUS	AY040839	3959 bp	DNA	linear	PLN 01-APR-2002		
DEFINITION	Aspergillus niger endoglucanase C (eglC) gene, complete cds.						
ACCESSION	AY040839						
VERSION	AY040839.1	GI:19879405					
KEYWORDS							
ORGANISM	Aspergillus niger.						
REFERENCE	Aspergillus niger						
AUTHORS	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus. 1 (bases 1 to 3959)						
	Hasper,A.A., Dekkers,E., van Mil,M., van de Vondervoort,P.J. and de Graaff,L.H.						
	EglC, a new endoglucanase from Aspergillus niger with major activity towards xyloglucan						
	Appl. Environl. Microbiol. 68 (4), 1556-1560 (2002)						
	11916668						
	2 (bases 1 to 3959)						
	Hasper,A.A., Dekkers,E. and de Graaff,L.H.						
	Direct Submission						
	Submitted (18-JUN-2001) Molecular Genetics of Industrial Micro-organisms, Wageningen University, Dreijenlaan 2, Wageningen 6703HA, The Netherlands						
	Location/Qualifiers						
	1..3959						
	/organism="Aspergillus niger"						
	/db_xref="taxon:5061"						
	<1038..>3841						
	/gene="eglC"						
	join(<1038..1137,1185..1329,1373..1571,1618..1858,1908..1961,2007..>3841)						
	/gene="eglC"						
	/product="endoglucanase C"						
	join(1038..1137,1185..1329,1373..1571,1618..1858,1908..1961,2007..3841)						
	/gene="eglC"						
	/codon_start=1						
	/product="endoglucanase C"						
	/protein_id="AAK77227.1"						
	/db_xref="GI:19879406"						
	/translation="MRRTPALVAAGTLLDSVHAASQATYTKNVYTGCGGTFPGIYF						
	NPSKAGVAARTDVGATRLSDSDTWTPLMDWANNSSWHDGIDAIVDPVDIVYV						
	AVGWYNDMDPNDISILRSTDGDTWEETKLPKFGVGNMGKRGHERLADPNDSIL						
	YFGVRSNGNLKSTYDGETWSNVAFAKWTGTFQDSSSTYSDPVGIAWTFDTSGS						
	SSGTPPLFVGVDVTGSEVFSVEDAGEETMWSPEPMYGLPHKGIISPSEHTLYISY						
	NSAGQDTGNGVTHKKNYKNTSGVMTDIPSMSDPTFYCGXGLADLVDPCTVVAALN						
	CMWDELINRSTDSGGSPTPMANGTIPSINLYIYSDISNAPMLDDDTSDPEVRVG						
	MMVVALAIDPDDSHMLYGTETLYTGGDILQNMSENNVTIESLAVGLEMAVGLIT						
	PPGPAALISAVDGDGCFHTSLTPAPQYHTPTVYSTNGIDVGNKPAIVRSGSD						
	SDPLTALSSFGESMVAIDYASSSTATGOVALSDADPTILMNSDGAIRANSATLSA						
	VSSLSPGAVIASDKANNTEFYGAGSGSEFLVSDTAATFTVTTIGSSTANVAIPADS						
	LAGWVYSTDNGLEFHSYNGKSPFQISGCTEGHSFECFKRPSDGDYPLYEGFTYDVG						
	VTGLFKTEDQVNMQIISDAHEGFGSASNAVNVNDLQYGRVEGTNGRGIYFGDPSSG						
	TLPEPATVASSASTAVKSSSTSTSTSVKVGSSSTVSSSTATVITTSLSKSTLTLLTWS						
	SSSTSTSTSTATCTAASAYCGCGSGSGFGTPQCPGAMCTFENEYYSQCKSIPIGIATDR						
	G"						
BASE COUNT	898 a	1110 c	1000 g	946 t	5 others		
ORIGIN							
Query Match	12.3%	Score 334.6	DB 8	Length 3959			
Best Local Similarity	53.0%	Pred. No. 5.8e-44					
Matches 1096	Conservative	0	Mismatches 754	Indels 219	Gaps 9		
0Y	128	GCGCGCGCGCTTGCTGCCCGGATCATCTTCCATCCCAAGACAAAGCGGTACATAT	187				

Db	1134	GGCGGCGAGGCTTCACGCCGGGATGCTTCAACCCCTTCAGCTAAAGGTGCGCATAT	1243
QY	188	GCACGACAGATATTTGGCGGGCTGTACCCGCTCAACGCCGAGACATCAGACCCCGCTC	247
Db	1244	GCTGCGACCGACATCCGGCGGGTGCATATCGGTTAACTCGGAGCATATCATGACCCCGTTG	1303
QY	248	ACGATGGGATTTGCTGATTAATGCCGCT-----	275
Db	1304	ATGCACTGGCGCAACAACCTAACTGGTACGTAAGCTTCCTACAGATAGCCTAGCCTC	1363
QY	276	-----GGCAACAATGGGGGCAATCCAGCGTGTGGCTGTGATCCGACAGAGATCAAA	327
Db	1364	ACTGTCCAGGACCATTTGGGGCAATGTAGTGATCGCGACCGACCCCGTTCGATACCGATCG	1423
QY	328	GGTGTATGCCCACTCGGCATGTATATACGAACACCTGGATTCGAGTAATGGAAGCCATAT	387
Db	1424	GGTATATGTTGCAATTTGGCATATGTAACCAACGACTGGATTCGAATGAGCGCTCATCT	1483
QY	388	TGCGCTCGTACAGCCGCGGCCAACGTGTCCTTTCACCAACTTGCCTTCAAGTGGGGG	447
Db	1484	GCGGTCGACCGCACCGGCGGCATACCTTGGAGAGAGACAAAGCTTCCCTTCAAGTTGGAGG	1543
QY	448	TAACATGCGAGGAGCGGAGCCGAG-----	473
Db	1544	GAATATGCCGGGTGGGGGTCTTGGCGAGGTAGTCTTATATCTGGGTTTGTCTAGCATAC	1603
QY	474	-----AAGCTGTGGCTGTGATCCGCGCAACTCCAACATCATCTTTGGTGTG	521
Db	1604	ACTAGCCGCTGCACACGCTGTGGCTGTGGATGCCAATGACAACAGCATCTGTACTTTGGTG	1665
QY	522	CTCGCTCAGAAACGGCCTCTGGAAGTCTACGAGCGCGCGCTGACCTTTTCCANGCTCT	581
Db	1664	CCCGAGTGGCAAAAGGCTTGTGGGAAAAAGCACGGACTACGCGACATGTGTCATATGTGA	1723
QY	582	CGTGTTACGGGCAACTGGGGCGTAATCCCAACCCGAGTGTATCCAAACGGCTCAACAA	641
Db	1724	CTGCTTTAACTGACCGCGCATTTATTTCCAAATTTCCAG-----CTGACCTTCACGT	1777
QY	642	GCACAACCAAGCACTCATGTGGGTTACGTTGCATCAACACAGACAGACCGGGGGAG	701
Db	1778	TGTGATCCGGTGGGAATTGCTGGGGTCACTTGACATCGAATCAGAGGTCCAGTGGCTCTGC	1837
QY	702	CCAGCTCTCGATATCTTTGTTGGCAGCGCTATA-----	734
Db	1838	CGACTCCGCGCATCTTGTGGGTATGAAAAATTTAGATATCTCAGACAGCCCACTG	1897
QY	735	-----ACATCACTGCTTCAGTATATGTGAGCAAGAACCGGCTCC	775
Db	1898	ACCATTACAGGTGTGTGGACACTGGAGAATCCGTTGTGTGCCAAGATCTGGCGAA	1953
QY	776	ACGT-----GGAGTCTGTA	790
Db	1958	ACTTGTAAAGCTTCTCCACGATCCGAATTAACCATCTTAACCAACTTAAAGCACTTGGGTA	2017
QY	791	CCGGGGCAGCCAGGAA--ATACTTCTCACAAGCGAAGAACTGCAGCCAGCAGAGAAG	847
Db	2018	TCTGGAGCCCATGTACGGCTTCTCTCCCAATAAAGGATCCTGTGCGCTTCTAGACAC	2077
QY	848	GCCTTGATCTGACCTATTTCCGATGGCAGACAGGGCCGTATGATGGCACATTGGCTCAGTG	907
Db	2078	ACACTTACATTTCTCTCACTCAACGGGCGCGGTCACTAAGATGGCAACATATGGACCTGC	2137
QY	908	TGGAGTATCGCAATTTGCAGGGGGAAGCTTGGAAAGATCAACCCCTGTCTGATCAAT	967
Db	2138	CACAGTACAAACATCACACAGCGGCTGTGAGCCGACATATGCTCCACCTCCATGACACAC	2197
QY	968	CTATACTTTGGCTTTGGCGGCTTGGCTTCGCTCGATTTGCCAAAAGCAGAAACCCCTTGTGTT	1027
Db	2198	ACCTAATATGCTATGGCGGAGCTGGAGCATCTCCAAAGTCCCGGGAACCTCATGTGTA	2257
QY	1028	GCTTCTTTGAACCTTGGTGGCCAGATGCTCAAGCTTTTGGTGCAGCGACTCTGGGACA	1087

```

Db 2258 GCAGCACTCAACTGCTGTGGGCTGATGAGTGTGGCGGACAGCATCTTGGGGG 2317
Oy 1088 ACATGAGCCGATCTGGGCGGCGGATATCCGATGAGATCTATTAACAGCATC 1147
Db 2318 ACCTGGTACACCTTGTGGGTTTGGAGGCTATCCAGCATCACTACTACTCTTAT 2377
Oy 1148 TCAACTCCCAAGACCGGTGATCAAGAACACTTATGATGTGACGAGGAGTCAACG 1207
Db 2378 GACATCTCCAGGCCCATGGTACAAAGA-----CGACACATCT 2416
Oy 1208 TCCGATGATCTCAAGAGCCCTCGGTGATGATGATGATCTCTCGAGATTTGACCC 1267
Db 2417 ACAGAGGAATTTCCGTCGCTGCGGTGATGATGATGATGATGATGATGATGATGAT 2476
Oy 1268 GACAGCAACCACTGCTGCTACGCGACCGAATGATCTTTGGGCGGCGAGATCTCACC 1327
Db 2477 GACTCTGACACCTGCTTATGCGGACATGCGGAGACATCTACGCGGCGGCGGCGGCG 2536
Oy 1328 AACTGGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1387
Db 2537 AATTGGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2596
Oy 1388 TCCGTCACAGACTGCGCTGCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1447
Db 2597 GCTGCTTGGGCTCATCTCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2657
Oy 1448 GACAGCGGCTTCACTTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1507
Db 2657 GATGGGCGCTT-----CTACACACTTCTCTGACGCGGCGGCGGCGGCGGCGGCGG 2707
Oy 1508 GCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1567
Db 2708 CACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2767
Oy 1568 GTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1627
Db 2768 ATGTCGCGCTCTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2827
Oy 1628 GCGAGCTGAGCATGCACTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1687
Db 2828 GATCTGTTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2887
Oy 1688 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1747
Db 2888 TCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2944
Oy 1748 TTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1807
Db 2945 AACTCGGCTACTCTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3004
Oy 1808 AAGAAGCAACAGCGCTCTTCTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1867
Db 3005 AAGGTAACAACAGCTCTTCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3064
Oy 1868 ACCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1927
Db 3065 ACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3124
Oy 1928 GCTGCTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1987
Db 3125 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3184
Oy 1988 TCGCAGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2047
Db 3185 TCAACCACTACGCAAAATCTTACACCA 3213

```

```

RESULT 5
AF078038S1 3262 bp DNA linear BCT 11-FEB-2001
LOCUS AF078038S1
DEFINITION Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 5 gene, partial
cds
ACCESSION AF078038

```

```

VERSION AF078038.1 GI:12743874
KEYWORDS 1 of 5
SEGMENT Caldicellulosiruptor sp. Tok7B.1.
SOURCE Caldicellulosiruptor sp. Tok7B.1.
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales;
          Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE 1 (bases 1 to 3262)
AUTHORS Gibbs, M.D., Reeves, R.A., Farrington, G.K., Anderson, P.,
          Williams, D.P. and Bergquist, P.L.
TITLE Multidomain and multifunctional glycosyl hydrolases from the
JOURNAL extreme thermophile Caldicellulosiruptor isolate Tok7B.1
MEDLINE Curr. Microbiol. 40 (5), 333-340 (2000)
PUBMED 20171169
REFERENCE 10706665
AUTHORS Gibbs, M.D., Reeves, R.A., Farrington, G.K., Anderson, P.,
          Williams, D.P. and Bergquist, P.L.
TITLE Direct Substitution
JOURNAL Submitted (27-OCT-1999) Biological sciences, Macquarie University,
          Sydney, NSW 2109, Australia.
FEATURES
source Location/Qualifiers
1..3262
/organism="Caldicellulosiruptor sp. Tok7B.1"
/isolate="Tok7B.1"
/db_xref="taxon:80339"
274..>3262
/note="ORF5: multi-domain"
/codon_start=1
/transl_table=11
/product="glycosyl hydrolase 5"
/protein_id="AAK06388.1"
/db_xref="GI:12743879"
/translation="MKRLSLFLILFLIAEFLQAKADGTIAVOPVYKWKI
EGGGRTITGVNPKKNLYVVRTDGIYASRTDGNMTLMDVSPQDMNLGYES
IATDPVDPKXVILACOGITVNSMTDMNGILNLTSDGDFETPLPFXGKMPKXRL
GERLATDPNNRILILGTRGKGLMSDEGYSMKRVISFPNGTITIEDPNCNPDLN
HITGVVAVVDFPTSGRPGEGXIIYGVAVDKTTSIYTKDGGTQAOALDGPQGLLPQ
RAKLSDDMLYITYSNTQGVNGDGEVAVRYMTKGEKMNISPMADLPQFYGILAV
DAONPKVYVVAALSSWMDPTYIMRSIDGSEYKCIEMNGYPNRLTHYIMDISAAML
NFGXNBPPTPEVSPKLGWVGLTEIDPESDXMLVGTGXILVGCDDLTMWDKGNITTI
KKAICIEETSVQALISPRVGBLSALGDIAGFREDLEKAPNNYVNOYNNMCTTDDI
DRAELPNPMTAVRQVNDKOMPNTNRTIGFSYTGGSFMRPEGTSEGTAAAD
GSAVWAPKGVAVCYSTDGNKWEKANVSEAIYSDRVNPNKTYAFENGFYISAD
KCKTFIESPAALPISGNFKTVPIGEGIDIMLVNGMNMSTDGYSFVYISGVEDPAS
IGFKRPAEGEYPAIYTYAKINGVRGIFERSDDCKTWIIRINDKHQFCGANADITGP
RVYGRVETATNGIKWELIAYSNISPSATPTSPSPPTPTPTATITSTPAKPTSP
PPTPTPTPTATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
SAGSIRPWFKIYNGSSSYDLSRVKIRIYTYDGRKPSAVCDMAQIGASNTPTNFVK
LSSGVSGADYILEVGFSSGAGDLPKDKDGIQVPRNKNMDSNYNOADWSLQSWTD
YGENKAVTLVYGVGLVWGEPPGA"
274..2580
/note="Region: endoglucanase domain"
/evidence=experimental
misc-feature 2791..>3262
/note="type IIib cellulose binding site"
misc-feature 2581..2790
/note="PF-linker"
BASE COUNT 1056 a 529 c 794 g 871 t 12 others
ORIGIN
Query Match 9.0%; Score 243; DB 1; Length 3262;
Best Local Similarity 50.5%; Pred. No. 2..7e-29;
Matches 685; Conservative 8; Mismatches 643; Indels 21; Gaps 4;
Oy 107 TGGAAAGACTCAAGCTCGGCGGCGGCGGCGGCTTCCCGGCGGCAACTTCATCC 166
Db 388 TGGAAAGATGTAAATTAAGAGTGAGGAGGCTTTATCACTGGAATAGCTTAAACCG 447
Oy 167 AAGACAAAGAGCGTATGACGAGAAAGATTTGGCGGCGGCTGACCGCCCTCA--AC 223
Db 448 AAGAGAGAAATCTTGTTTACGTTAGACAGATATAGAGACCTTACCGCTATACAGAC 507

```


Oy 224 GCCGACGACTCATGACCGCCGTCACGATGGATTGCTGATTAATCCGCGTGCGACAAC 283
 Db 508 GGTGCGAATACATGACACAACTATGAGCTGGATTGATTTGATGATTAATTTGCG 567
 Oy 284 TGGGCGATCGACGCTTGCTGCTGATCCGACGACGATCAAAAGGTGATGCCAGTC 343
 Db 568 GGTGAGAGAGCATTCGACACATGCTGTTGATCTCAATCMGGTTATATGTCGCCAC 627
 Oy 344 GGCATGATACGACAGCGGGATCCGAGTATGGAGCATTCGCTCGCGACGCC 403
 Db 628 GGCATTTATACCAATAGCTGGACACATATGATGCTGCTATTTGGCTTCTGACGAA 687
 Oy 404 GCGGACGCTGCTCTTCAACCACTTGCCCTTCAAGTCGGGGTAAACATGCCAGAGCC 463
 Db 688 GGAGATACCTTTGAAATCACTCCACTCCCTTTATATTGGTGAAGAACATGCCRGAAG 747
 Oy 464 GGAGCGGAGACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 523
 Db 748 AATCTGGGAGAAAGGTTCATGATTCACCAATATATGATATATATGATGATGATGAT 807
 Oy 524 CGCTCAGGAAAGCGGCTCTGAGAGTCTACGAGCGGCGGCTGACCTTTTCAAGGCTCG 583
 Db 808 AGAGAGAGGAAATGGGTATATGAAAGTGAAGTATGAGTGTGATGAAAGAAAGTACAA 867
 Oy 584 TCGTTCACGCGCACTGGGACGATACATCCAGACCGGAGTATTCACAGCGCTACAAAC 643
 Db 868 AGTTTTCCTAACCCAGGACATATATTTGAAGATCCAAATGCCCAAGATTTATTTGAC 927
 Oy 644 GCAAGCAAGGACTCATGAGGCTTACGCTGCACTCAACGACGACGACGACGACGACG 703
 Db 928 CATATACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 987
 Oy 704 ACCTGCTGATCTTTGTTGGACGAGGCTGATATACATCACTGCTTACGCTATGAGACG 763
 Db 988 AACAAGATATTTATGCTGGGTGTTGAGATTAACCA---CAAGCATATATATACAAA 1044
 Oy 764 AATCCCGCTCCAGCTGAGTCTGATCCGCGGACGCGGAAATACTTCTCCACAG 823
 Db 1045 GACGGTGGCAGACATATGCGAACATTTCCAGACATTCAGGCTTACTCTCCAGCT 1104
 Oy 824 GCGAAGCTCAGCCAGACAGAGAGGCTGATCTGACCTATTCGATGGACAGAGGCG 883
 Db 1105 GCAAAATTAAGTCTGATGGATG---TTATATATATCTTATATATACACTGACGGG 1161
 Oy 884 TATGATGGACACTTGGCTGCTGAGTGGAGTACGATTCAGGGGAACTTGGAAAGAC 943
 Db 1162 TATATGATGATATGAGAAATGAGATATCAATATACGCGGAAATGGAATAAT 1221
 Oy 944 ATACCCCTGCTCTGAGTCAATCTATCTTTGCTTGGCGGCTTGCCCTCGATTTG 1003
 Db 1222 ATATGCTCCATGCTGCTCAAGATCTTATTTGGATATGCTGATGGCTGACGCC 1281
 Oy 1004 CAAAGCCGAAACCTTGTGCTGCTTGAATCTTGGGCGGACAGATGCTGACGCTG 1063
 Db 1282 CAGATTCGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1341
 Oy 1064 TTTGCTGACGACTGCTGGAGACATGAGACCGGATCTGGGCGTGGCGAGCTATTCG 1123
 Db 1342 TGGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1401
 Oy 1124 ACTGAGACTTATCTACGATCTCAACTCCCAAGACCGCTGATCAAGAACACTTT 1183
 Db 1402 AATTAACATTTGATATATATATATATATATATATATATATATATATATATATAT 1451
 Oy 1184 ATGATGATGACGAGGAGTCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1243
 Db 1452 -CTTTGGAATATCTATCTACACCTGGAAGTAAAGCCAAATTTGGCTGATGTT 1509
 Oy 1244 GAGTCTCTGAGATGACCAACGACGACGACGACGACGACGACGACGACGACGACGAC 1303
 Db 1510 GGAACACTGTAATGACCATTTATATGATATATATATATATATATATATATATATAT 1569
 Oy 1304 ATCTTTGGCGGCGAGATCTACCAACTGAGACCGCGCACATGCTGCAATCAATCA 1363

Db 1570 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1629
 Oy 1364 CTGGCAGACGCGATGAGGAATTTCCCTCCAGACGCTGCTGACACCGCGGAGAC 1423
 Db 1630 AAGGCTATATGGAATTCGAACTTCGACGACGATGATGATGATGATGATGATGAT 1689
 Oy 1424 GAGCTATTTGGCGGAGTGGAGACGACGACGACGACGACGACGACGACGACGAC 1460
 Db 1690 CATTTATTCAGTGCACCTGGAGATATAGCGGATTTCA 1726

RESULT 6
 AE007608
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AE007608 12732 bp DNA linear BCT 27-JUL-2001
 Clostridium acetobutylicum ATCC824 section 96 of 356 of the
 complete genome.
 AE007608 AE001437
 AE007608.1 GI:15023819

Clostridium acetobutylicum.
 Clostridium acetobutylicum.
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

1 (bases 1 to 12732)
 Nolling, J., Breton, G., Omelchenko, M. V., Markarova, K. S., Zeng, Q.,
 Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hitt, J., Wolf, Y. I.,
 Tatusov, R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P.,
 Daly, M. J., Bennett, G. N., Koonin, E. V. and Smith, D. R.
 Genome sequence and comparative analysis of the solvent-producing
 bacterium Clostridium acetobutylicum
 J. Bacteriol. 183 (16), 4823-4838 (2001)

2 (bases 1 to 12732)
 Childress, D., Zeng, Q. and Smith, D. R.
 Direct Submission
 Submitted (24-JUL-2001) GTC Sequencing Center Production,
 Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
 Beaver Street, Waltham, MA 02453-8443, USA

FEATURES
 source
 1. 12732
 /organism="Clostridium acetobutylicum"
 /strain="ATCC 824"
 /db_xref="ATCC:824"
 /db_xref="taxon:1488"
 73. 2592
 /gene="CAC0919"
 73. 2592
 /gene="CAC0919"
 /codon_start=1
 /transl_table=11
 /product="probably secreted sialidase; several ASP-boxes
 and dockerin domain"
 /protein_id="AAK78895.1"
 /db_xref="GI:15023820"
 /translation="MNKRIVSMVAGLSIFFTGEVTHHSAAANKAAASVOSQYKMDN
 AKIGGVPAVIEKTEKDLIYARTDGGRAVBMKANKNRPITDGSMDMLGCS
 IATDIDPNRYIAAGLITNDMODNATILSSQDKGNWKRQLPFKGGNNPGRNMG
 ERLOIDPNDDKILYUGARSNGMLKMSBDTGOWSVDFPDGIVDQUPDNEYLADKY
 GVWETFPDPSGTGKSPQTYVGAADTGNNTYVNDGRTWSAVKQDPKGYLPHHG
 ILASDGLYISYNSCGPYDSGQVWYNTKRGEMNITPPAVDTRSGRGLSVDA
 ONPNVVAATINRMPPDEIYRSDPAGTKMPIMDMNGYPNRTLCYNLDYSAQPLDM
 GKTCVTPDPLVTKLGLMMGDLIDPFNSDRMFYGTGATLYGDDLTMDKKNDISY
 KANGIEECANVDVYPTKGAOLLSVSDGCEYHDDITKVSQKMTTPNFSAITSDY
 AESVFNVEVRGNVDTSKNOQDKDGISTDGKMFSGSNTSGYTKAGTYAAGADAK
 TIVSPPEGANAAYSTDNKNWTPCSGVPQAKVSDVNRKPEYGLNGKFLYSTDA
 GATFQSSQGTGLPTKSGKIFKTVIGHEGDIWLAGSKDGLMSTDSGATFTYVSGVDS
 DTVGKSKTDDGPALYIMDATIDGTATIFSDSDGATVWRIINDAHQYSPDYCTIG
 DPNKGRVETGNGRGIYDIDSGQPPPTSPPTSGDGVNDGDIINGRWVVLTVYIAG
 KDVIIDKRAADYVNGDGVNGRDMLRLRYISGKITPSBSTK"
 complement(2634..2978)
 /gene="CAC0920"

gene

CDS complement(2634..2978)
/gene="CAC0920"
/codon_start=1
/transl_table=11
/product="Protein related to MRP/DOPD protein family,
function in bacteria is unknown"
/protein_id="AAK78896.1"
/db_xref="GI:15023821"
/translation="MPINSTVTVKLDEKEESTLKVLELGIITELPGKSEDLWVGFK
DNPLVFGKEKKDKAPFIEVKIFGSADASKNKLVAISLSIMENELSIKDCIYITVD
EVSNNMNGSLEF"
gene 3104..4252
/gene="CAC0921"
3104..4252
/gene="CAC0921"
/codon_start=1
/transl_table=11
/product="Metal-dependent amlodihydrolyase"
/protein_id="AAK78897.1"
/db_xref="GI:15023822"
/translation="MTTEKOEFLKLKLGVRRELHKHPELSMKEVETTKRIKRLYE
NDIKVLDPEVGVVAIEEGHEGKTVAIADIDALPIEEKTELPSSVNNKIMHACG
HDFHTAALIGTAIIILSKRAEITGRVRIIFORGEFGKSGOYIEEGVLRDVLCEFG
HNKPDLPVCTGIRSGELMASVDREIKVIGSGAGACIPNKCIDPIVVASQIVASMG
VVSRSISPIESVVISITRPNSTGMVNSDSKEMEGVKTFFNNITBEIRNLMEHTAK
GIARALGGKIEFKWFSYSPLENDVRFKDLCEAEDNGKYVEAKRNLDGGEDFSFYQ
TVPSFFVMVGDTGEEMHKANYNLKEDLWVANYFSDLALKVLKSK"
gene 4594..4863
/gene="CAC0922"
4594..4863
/gene="CAC0922"
/codon_start=1
/transl_table=11
/product="Uncharacterized small conserved protein, homolog
of yfjA/yukE B. subtilis"
/protein_id="AAK78898.1"
/db_xref="GI:15023823"
/translation="MFDRALCEDEIKKSTIEKSTQOLDSWGESCTAFETKYLILDR
KNTYTEILOEYSKALYDISETYONTDDIANKAVNAEASEOK"
gene 5056..6348
/gene="CAC0923"
5056..6348
/gene="CAC0923"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein; CF_1 family"
/protein_id="AAK78899.1"
/db_xref="GI:15023824"
/translation="MOSLKVDMRNGLLKINEAKOGKSAVDKESLKNKRSRSTIEELKN
LKRELKAVNOLEGEIRKINSNDYVYKRTDVKCASLKTSTYTRKRYVLLNFG
EVERPVIGIGIVPIDAVESWYDRNKFVLKTTITLLEGAIGCVITLPEAGITWKA
WDLSAAGVMSLDNI GTAASAIYQKRYEKSEDSGDLFOSIFGGAGWSARILG
VNPSPYSCSSGKTLVDRTSATVTVLVNGVSYSLCKDVEKLPALKAEASSNKTLEA
ANNAEKIHMDAMATKRNKINGINNAKKEVYDALKKKNVGTNPFEPYKQPFKKDKML
KKDGKMLPOLQODLLIODDLETFEPKNTLNAKNKNSIRDLNKNYRDIKASLVNIPA
GGTLNTEDVOSFHFNFPEBIRKIFPEIKVVLGGER"
gene 6345..6923
/gene="CAC0924"
6345..6923
/gene="CAC0924"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein"
/protein_id="AAK78900.1"
/db_xref="GI:15023825"
/translation="MIRFOKRAIAFFLTISILICODEKIIISFGIEFDILCNLFPI
GGIIGFRIYRWYKROIKKYTELKEDRLDKLEYIYDERGVGVYNAFVINDGVC
KIRFPOYCKDNDNFEYLEPIYKKNGEKFDIAVADKNDYKKSIESINTECEBEK
AEKRRKRRLNFTVNSRMELAKKYEBAKNNK"
gene 7003..8043
/gene="CAC0925"
7003..8043
/gene="CAC0925"

/codon_start=1
/transl_table=11
/product="TPR-repeat-containing protein"
/protein_id="AAK78901.1"
/db_xref="GI:15023826"
/translation="MENENVLRQAOLNNRGLIFIRNAGDEALKFNAIKEDENFKE
ATLNKAALCELANNVDAVASCYDKLITIRYPNGKAYFGKGVWFYKONIKRAILEYN
RSIYLGESEGVYVNLALCMESIGELDIAIKWLNRAILINDONIITLNKKAVALYIK
KFEALDECYNRYLKEVDNEEYVHFKAMLGEMGIERAFETISRGELALGCOMTLCY
DKVAFPEKOFERKALECTEKALFLDEGVLLILKKGDLTYLKKIDEAKSYYDKILE
PHENMEGVFAKANCMLGHEHDAEFFECTEIERLSGSEESVLINSYYRRALNLKKA
KTRQKRILIRKO"
gene 7998..8357
/gene="CAC0926"
7998..8357
/gene="CAC0926"
/codon_start=1
/transl_table=11
/product="TPR-repeat-containing protein"
/protein_id="AAK78902.1"
/db_xref="GI:15023827"
/translation="MGNKEAKVAKKAIKYNFLIKYPYDVKLNFPKANCLRDIED
VEKAEILEYITLDLNDALLEVYLMARNRISLNKYEAKKEDLNTVYIKLNPAVKYIEL
DQILKEPISGSESCYR"
gene 8379..8673
/gene="CAC0927"
8379..8673
/gene="CAC0927"
/codon_start=1
/transl_table=11
/product="Hypothetical protein, CF_2 family"
/protein_id="AAK78903.1"
/db_xref="GI:15023828"
/translation="MNSDEIRIDERACDVIKDLKAVGEKVEFYNLILHSNRHIEDG
IKGSADVSTTETNKYIKVYKEMQNDVAANNVYSFPAEDKMKIKTKYR"
gene 8713..9012
/gene="CAC0928"
8713..9012
/gene="CAC0928"
/codon_start=1
/transl_table=11
/product="Uncharacterized small conserved protein, homolog
of yfjA/yukE B. subtilis"
/protein_id="AAK78904.1"
/db_xref="GI:15023829"
/translation="MVRCDIEIDASIFEETIKVYSDSKVRHLIDLCLENELKRMED
TWECDAKKEFDSTFPGFYSAMKDCDMLLEILKEILKLVTSFESIDTEMKNDKK"
gene 9712..10335
/gene="CAC0929"
9712..10335
/gene="CAC0929"
/codon_start=1
/transl_table=11
/product="SAM-dependent methyltransferase"
/protein_id="AAK78905.1"
/db_xref="GI:15023830"
/translation="MGSMDYFDNINAASNNVIRSEFEERLKKYVSKYSIKRKTADF
GGCTGFTSLASDANWVFGLDNSNNMELKLSHDKRLNINYLKSSLSVTLFDE
SIDAVEINMALHIKIDAEKAIKEMVRLKDDGTIVISDSEHNGWAREMEDWELGF
SNEQIDIMWLKKNVFNIDIENTDLCKCKGSSRGEHTETSIYAKAAK"
gene 10353..11498
/gene="CAC0930"
10353..11498
/note="metB"
10353..11498
/gene="CAC0930"
/codon_start=1
/transl_table=11
/product="Cystathionine gamma-synthase"

Query Match 7.5%; Score 203.6; DB 1; Length 12732;
Best local Similarity 50.0%; Pred. No. 4.6e-23;
Matches 667; Conservative 0; Mismatches 634; Indels 33; Gaps 5;

OY	128	GGCGGGCGGCGCTTCGCGCCCGGCATCATCTTCCATCCCAAGCAAGAGCGCTACAGATAT	187
Db	214	GGTCTCGTGGATATGTTCCAGCAGCTATATTATTAATAAACAAGAAAAAGATTATATAT	273
OY	188	GCACGAACAGATATTTGGCGGGCTGTGA---CCGCTCAAGCCGACAGCATCATGAGCCGC	244
Db	274	GCTGTACAGATATGAGGAGGACGATATAGATGGGACAAAGGCTAATTAATGAAGTGATCCCA	333
OY	245	GTCACGGATGGGATTCGTGTAATATGGCGGTGGCACACTGGGGCATATGCACCTGTTCGG	304
Db	334	ATTACTGATGGATTTCTGTAT-----TGACAACTCTTGGCTGGCGAAGATATAGCA	384
OY	305	CTTATCCGCGAGCAGATCAAAAGGTGTATCCGCGAGTGGCATGTATACGAACAGCTTG	364
Db	385	ACTGACCCCAATAGACACTAATTCGTGTTTATATAGCAGCAGGTTTGTACCAAAATGATTTGG	444
OY	365	GATCCGAGTAAATGAGCCATCATTTCCGCTGCAGACCGGGCGCAACGTGTCCTTCAAC	424
Db	445	CAAGATGAATAATGCATATATCTTATGTTCCACAGATTAAGGTAAACATCATGGAAGAATAC	504
OY	425	AACCTGCCCTTCAAAATGCGGGGGTAACATCCAGAGCGGGAGCGGACAGAGGTGCGGT	484
Db	505	CAACTTCATTTAAGTAGAGGATTAATATCCGGGAGAAATATGGGAGAAAGACTTCGA	564
OY	485	GTCGATCCGGCAACTCCAAATCATCTACTTTGTGCTCGCTCAGAGAAACGGCCTTCG	544
Db	565	ATGATATCCTAATGATGATTAATAATCTTATCTTGGAGCTAGAAGGTGTAATGAGACTTTGG	624
OY	545	AAGCTTACGAGCGGGCGGCGTACCTTTCCAAAGTCTCGTCTGTTCAACGCGCACTGGAGC	604
Db	625	AAAAGTGAAGCTATGTGCAACAACATGCTCTAAGGTTGATTAATTTCCCTCATACAGCGAC	684
OY	605	TACATCCAGACCCGAGTGTATCCAAAGCGCTACACAGCGACAGCAAGAGACTCATGTGG	664
Db	685	TATGTACAAAGATCC-----TCAAAATGAATATACAGAGATTAAGTATGAGTAGTTGG	738
OY	665	GTTACGTTCCGATCTCAACACAGCAGACCGGGGGAGCCACGTCCTGTATCTTTGTTGGC	724
Db	739	GAAACTTTGATCTCTTCAACAGAGAAACAAAGGAGCCCAACTCAATCTATGTAATGTGGT	798
OY	725	ACGGCTGTAAATCATCACTGCTTCAGTCTATAGTAGACAAATATGCGGCTCCAGCGAGAT	784
Db	799	GCACAGATTAAGACAGGAACAATATTTACGTTACTAATATGATGACAGAAAGCATGTAGAT	858
OY	785	GCTGTACCGGGGACGCCAGGAATATCTTCTCACAAGGCGAAACTGCAGCCACAGAG	844
Db	859	GCAGTTAAAGAGCAGCCAAAGGATATCTTCCACATCATGTGTAATATCTTCCATGTGT	918
OY	845	AAGGCTGTGTATCGACCTATTCGATATGGCAGGGGCGATATGATGGCACTTGGCTCA	904
Db	919	ATG---CTTTATATATATATATAGCAATACTTGGGAGCCATATGACGGCTCAATGTGACAA	975
OY	905	GTTGTGAGAGTACGACATTTGCAGGGGGAATCTTGAAAGACATACCCCTTCCTGTGATCA	964
Db	976	GTTTGGAAATATATATACCAAAACAGGAGATGGAACAACATTAATCTCACCTGCAGTAGGT	1035
OY	965	GATCTATACTTGGCTTTGGCGGCTTGGCCTTCGATTTGCAAAAGCAGAGAACCTTGT	1022
Db	1036	GACACTAAATCTGGTTTTTGGAGGAATATCGTAGATGACCAAAATCAATATATGTTGA	1095
OY	1025	GTTCTCTTCTTGAACCTTGTGGTGCGCAGATGTACAGCTGTTTCGGTGCAGCCAGCTGTGG	1084
Db	1096	GTTCCAACTTAAATTAATATATGTTGGCGGAGTAGAAGAAATATACGAAGTACATATGACAGT	1155
OY	1085	ACAACATGAGACCCGATCTGGGCTGGGCGAGCTATTCGACTGAGACCTATTACTACAGC	1144
Db	1156	AAGACCTGGAACCTATATATGAGATTTGGAATGATATCCAAACGTCCTTGGGATATAC	1215
OY	1145	ATTCACACTCCCAAGAGACCGTGGATCAAGAACACTTATATGATGTGACGCGAGTCA	1204
Db	1216	TTACATTACTACGACAAACCTTGG-----CTTGATTTGGGGTAAAGCCGGGGTA	1266
OY	1205	CCGTCACATGCTCATCAAGCGCTTCGCGCTGAGTATGATGCTCTCGAGATTGACCCA	1266

Db	1264	ACACGACCAAGATCCATTAGTTAAATTTAGGCTGGATGGGAGACTTGAATTAATGATCCA	1323
Qy	1265	ACCGGACAGCAACCACTGGCTCTACGGCACCGGAATGACAAATTTTGGGGGCCACGATCTC	1324
Db	1324	TTTAACTCGGATGAGATGTTCTTACGGTACAGGTGCGCACTCTTTATGAACTGATGATATTTG	1383
Qy	1325	ACCAACTGGGACACGCGGCACAAATGTGTCAATTCACAACTGCGAGACGGCATCGAGGAA	1384
Db	1384	ACTAATTGGGATTAAGGTAGAGATGTAGACATCTCTGTAAGGCGAAATGGAAATTAAGAA	1443
Qy	1385	TTTCTCGTCCAGGACCTGTGGCTCTGCACCCGGCGGAAGGAGACTTATGTGGCCGCAATCGGA	1444
Db	1444	TGTGCGATTAATGATGTTGTGTAGTACCACCAAAAGGCGCACAACCTTTTAAGCCCTGTGGGA	1503
Qy	1445	GACGACAAACGGCTT	1458
Db	1504	GATGACTCGGATTT	1517

[illegible]

[illegible]

Dn	1695	CGCCACGACCAGCGCCGCACACGGCACCACCAACCGCACACGACCGCACACCGCACCGCACAC	1694
Oy	2272	GGAAGAAGCTGGGGGGGGGACACGGGGGGATTTCTGTGTGACAAGACAGACACGACAG	2331
Dd	1695	CACGCCCGCCACACACACACCGCCACCGCCACACACCGCCACACACCGCCACACCGCCAC	1744
Oy	2332	TACGTCCTTCCCACGTCGACACGACTGAGTGCGAGCGATTGTATTCACAGACCGGGC	2391
Dd	1745	CACCAACCGCCACACACACCGCCACCGCCACCGCCACACACCGCCACACCGCCACCGCCG	1804
Oy	2392	TTGCAAGGTGACTTGTGTGAGAGAACGCTGCGCGCGCGGTCCAC	2436
Dd	1805	CTCCACCGCCACACACACACCGCCACCGCATCTCGCCACCGCCAC	1849

LOCUS	AE012276	RESULT 8
DEFINITION	Xanthomonas campestris pv. campestris str.	BCT 23-MAY-2002
FEATURES	of 460 of the complete genome.	
KEYWORDS	AE012276 AE008922	
SOURCE ORGANISM	AE012276.1 GI:21112840	
REFERENCE AUTHORS	Xanthomonas campestris pv. campestris str. ATCC 33913. Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xanthomonas. 1 (bases 1 to 11910) da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluyt,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J., Chambergro,E., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locall,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Medanis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,I.A., Sena,J.A.D., Silva,C. de Souza,R.F., Spíndola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P. Comparison of the genomes of two Xanthomonas pathogens with differing host specificities Nature 417 (6887), 459-463 (2002)	
JOURNAL MEDLINE PUBMED	Nature 417 (6887), 459-463 (2002) 22022145 12024217	
AUTHORS REFERENCE	2 (bases 1 to 11910) da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluyt,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J., Chambergro,E., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locall,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Medanis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,I.A., Sena,J.A.D., Silva,C. de Souza,R.F., Spíndola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P. Comparison of the genomes of two Xanthomonas pathogens with differing host specificities Nature 417 (6887), 459-463 (2002)	
TITLE JOURNAL FEATURES		

Title Direct Submission Submitted (28-NOV-2001) Departamento de Biologia, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil

Location/Organism "Xanthomonas campestris pv. campestris str. ATCC I_11910 /organism"

33913"
/strain="ATCC 33913"
/db_xref="ATCC:33913"
/note="pathovar: campestris"
/gene="gppr"
/note="XCC1748"
241..1212
/gene="gppr"
/note="XCC1748"
241..1212
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="galactose-binding protein regulator"
/protein_id="AAM41039.1"
/db_xref="GI:21112841"
/translation="MANSGTPMFMNAKLTROLLLLHLHEQSRVLAEAASMTOPA
ASKLAEMEDMGVLFERNHARGVEPTMGHVLIRARAMEIGRAOELALRAGR
MCOASIGIVVNGTILVPOALAEVKEFPDILVRVEMOYRPLVAKLIDGODIYGR
ILGPEVGELEFEPLADEBPSIARAGHPLASRANLOHDLVRHGIILPPADSLAR
LDSPMEHGVOTPTNAIETSSLPTVSTILRGSDMLALPVEVAPLLOAKLLTVLPIE
LGRMSESGIIRRDVILPPGAEIRIOLARSTARRLYRGLRSPESPV"
1470..4382
/gene="fhua"
/note="XCC1749"
1470..4382
/gene="fhua"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AAM41040.1"
/db_xref="GI:21112842"
/translation="MTTSSHSASGAGARARLYRHRHSAIISTICVLLAARAYAOA
PLTNGGSAVDLDIVEVKRVASLIKQVIRKDSQIVDSVAEDIGALPDSVETL
PQVSGVTTIDREMARSDPDHFSAGSGVMIRGLTVNGELNGRDVESAAGKSEEDY
PALMAGVYVKNPSAEIIEGLGTVNLTMRPMPNPGRIVGANDVYVGMKSKFY
PASFLSDRMWTEIGEMFMIDIAHSELATSDGLOVEPVYRTDDVLAAGSREY
VPGGVNMRLOLFEIRTRDGIIAAFQKPSDAETIYAOLRSRYEMMORRAFPFNSN
NSTTPARCTTPYDERGATLRGAPVSSMRGALTGCVRFNTDNRISEORTTTTDSV
GSHFFNDMLQIGDMQLOVSENEOLDFTVFSAIYPLGLTVDSGYPSVQIANPFT
DQPSNFEAAMADHLGKNGRQLSTRVDLEFFDSSMLFARFGMARADRNQVANS
YNNQVISDNMAIIEGTNSGLNAMDPTVOPDNTROEPTAOAAVLAALFENDAMGIPV
VNNYGSASOMIOIIVLRGSGNAPDPTVOPDNTROEPTAOAAVLAALFENDAMGIPV
DQNGVRYVOTQOABRGQONLAGIANSPELOARYGQVFDNNSGCVTVVLEISN
LTLRSDSLQMRFAASKARPDYTOLOPYVLLNETNDAGCATDFVAGNPNLKP
VANOEDTLEWTFDTSMLYTLTFKRVADYFSTQRSIYDNRPMLVRYRPNMDEGT
LGAERGYQFEDSLPGMLNGFVANAFVDSKGGANATDPTWTYVYGVELPLEG
LSRRSYNLAGIIEKGPLRLRAYNMRSRLTTSDAATRLPTMADDDYGGDLSFFYRF
NHLOLQLOANNLTNTVTKVLMGPTSYEGGEVDPRLYTRSSFYVNDRRYSLVLRMN"
4555..6747
/gene="citra"
/note="XCC1750"
4555..6747
/gene="citra"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AAM41041.1"
/db_xref="GI:21112843"
/translation="MLPQGLKPCMDVHASMSPRLPRPDVAAATTTGGCVARAROK
ISTCYLGSTHSVQHRHPAGRAVQGRSIRPRRSYVAISVATLLSAQYKODATTT
AAPASDATTQULDTVOVIGTRSSVTKAQVLKQNAEIVDSIAEDGKLPDNNVAVL
ORI SGIQISRNNGEGSSIAIRGLTVRTELRNIDFTANDGSLSEDEVSABLLAEAL
VYKNSADMIIEGLGTVDLRTLRPLFDYGRKIASGVQYNYDLADGAPAFSGLFSD
RMOTGIGEIGILVNSOQSPRODITISLEPYRTODDLEGGOGSVPHGAGINTTI
GRRERTGAFAMQWRPNDSVEYTVOLRSDYPSMODYSFPALTOQNPLOGLPGIOVN
DNREYNGSFQVNPPTESNLSLTERHSVTTTDSIGAKWYNEKLLTSTDFDYDATTG
TMYIVATGODISPOFVNRGDLPRLSVTDSAGAEGLADVANNYDWRHRLDNKDK

GTEFAMRADMLAFDSDFYRSFKAGVRYTRDRAETKGNITRFMCINTACGATDSFPA
APPSIGIVNPTITDEFGRGESSYFGRLLTPANASVANSYEDTLATFGASPLVFAPNINA
OREKTYAAGVLRPGVDSDSTPDCDNGVRRVTRTEGSGVGTGDAEGCGILPVAQAO
QRTYDVLPSLNRHMLSDLOLRFAASRQISRPTRDLRPNLSLSTG"
6853..7455
/gene="XCC1751"
6853..7455
/gene="XCC1751"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AAM41042.1"
/db_xref="GI:21112844"
/translation="MMYGLTFYKKVEGFIAVFNENVYDGOVWOITRPYNGACKIRG
AELGYTFDFDLPGMISGGLDTNTTYTDSERPSPATDINOSLTVLEGLSKSYN
ALISYETARPOGRVAVNMRSMDLVTTAGGTGNLTPVYNGFGQLOLSLRFNTDWSI
SLDGVNLDTRRESYLGTESSRYRDEPVINDRYGLTLRSL"
7567..9822
/gene="XCC1752"
7567..9822
/gene="XCC1752"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="AAM41043.1"
/db_xref="GI:21112845"
/translation="MIMMIDPTKRRRTTTGAMLAVALSLLLLFSTPNVRAEPATSG
PYMRSVAIGGGFVTVGLFHPAEKRLAARVDVGAYFMDAQOQWTLTMDLADGD
NMKMGIDAEVADPADADALYLAAGTYMHERAGNAVALRFRNGRTERADLFFKGN
OLGRANGERLAVDPDGRVLLGSRDAGLMSRDDGAHAKVASFDAALACATARNH
VGRBOAGTAFVPPVPAASGNTGTPPIRYVYGTQSTSYVSEDDGSRSAACQPRG
LRPSHAGSGDHWLSTGDDQPGPDLMAAGALMKTPAGCRNRETSIPLOPAGSGRG
KCAVAVDPQOVLLASTFRRTPRDELRSYVDGGHMAPLADAVFODISAPWPAHA
PTHMMQALIDPFDGNHAEVTVGYGIMASRNLODRAAPORPLQNMFOEDGLEETVAD
LISPMAGALLLALDIDGFRHDELDRAQLOVAGRLTNGESIDAAGQAPQWVNSGT
VDRNRNETRALYSRFDGKQWAPFASEPAGGAGSITAGDAQAVVMAVERGMWRT
SPFGQNMOWDOLPTPTAVMADRVAAPVAVDVGASVLAAGLSESTDAARSPATGVOYS
PARDETRQLPDPDRAGVYTLASFGKVMKQOGTLOVLSQPDARSIGTGKILRA
GAPPLVLAGRGVGVFRSDGGVQWOMRINDAHFRGRPYSVTGDPRIAGRVFAT
GGGGLFYGDPR"
9926..11839
/gene="XCC1753"
9926..11839
/gene="XCC1753"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="sialic acid-specific 3-O-acetyl esterase"
/protein_id="AAM41044.1"
/db_xref="GI:21112846"
/translation="MTPSHISRACLGVLSSGSNAAPSLPULFADGAVLQREAMP
VGMATPGASIVSEFDGKOATKATDAOGCKWALPHATRFKGYVSVSHDGGALQVD
VLVGVVWLASGOSNMENMPLAOSDQPOAAANADLHRTFKVPSKMSVSGPEKLGROD
WKAAMPENAGASAVGYFFAKELASTGPTICITIVSTWGSIAEMMDASIGLKAEO
NOGATEIAIKORDAAAOATGKRIGRKPVADGTPQWRTEFDRDMDSIPIVQOVELS
GEGMDGIAWYRTFTFLTAEAKAAGVQLOVQIDSDITYVNGQVQGGQLEKQWNLPRV
YAVPAALNAGVNOJIAVRVEDLSGGGGMGPDEORFVQIGTORALAGKFPAPQVRV
SLTNDKNOLEPTLLVYNOIMHPILOPFVKGIWVQGETNATDCAVYRDOFAMIOWR
ABRGAKRLDPLFWOLANFRAGGDTGDLSPWALLRSOSKTLALPANTGOAVITIDGNPT
DIHPNKRBDVGRRLAARHVAVYGETLVYSAVFAFASADQASVQOEDLOSALQVR
GGGTLQGRVAVAGARFHPHQAQRILAGDRVIVASDAVAAPVAVTCMSENPADANLVR
EQLPVSPFTDWM"
BASE COUNT 2093 a 3908 c 3912 g 1997 t
Query Match 4.3% Score 116.8: DB 1: Length 11910:
Best Local Similarity 50.3%: Pred. 0.3: 5e-09:
Matches 452: Conservative 0: Mismatches 427: Indels 19: Gaps 6:

QY	128	GGCGGGGGGGCTTCGTCCCGGCAATCTTCTCCACGAAAGCAAAAGGCGTACCAT	187
Db	7726	GGCGGTGGCGGATTTGTACACCGGTGCTGTTTCAATCCCGGCAAGCTGTCTGGCCAT	7785
QY	188	GCAGCAACAGATATTGGCGGGCTGTACCGGCTCAAGCGCCGACGACTCATGACCGCGTC	247
Db	7786	GGCGGCACCGATGTGGGTGGCGGGCTACCGGCTGGGATGGCGAGGCCACAGTGGACCGCG	7845
QY	248	ACGGATGGGATTTGCTGATAATGCGGGCTGGGCAACCTGGGGCATGAGCGTGTCCGGCT	307
Db	7846	CTGACCGACATGTTGGGGCGCCGACGACTGAGAACCTGATGGGCAATGACGCAATTCGCGCTC	7905
QY	308	GATCCGACGAGCAGATCAAAAAGGTGTATGCGCGCATGTGCGCATGTATAGCAACAGCTGGGAT	367
Db	7906	GACCCCGCGGATGCGGATGCGCTGTATCTGTGGCGGGCGGACACTTATATGACAGAACGCG--	7963
QY	368	CCGANTATTGAGACCATCATTTGCTGCTGTCACAGCCCGGCGCAACGTGGTCTCTTCACCAAC	427
Db	7964	-CCGGCAATGGCGCGAGTCTGTGGCTGTTCACCCCGGCGCGCATTTGAGGCGCCCGAC	8022
QY	428	TTGCGCTTCAAAAGTGTGGGGGTAACTGCGACGAGCGGAGCGGAGAGCGTCTGGCTGTC	487
Db	8023	CTGCGCTTCAAACTGGGTGTATCCACGCTGGGCGCGCGCATATGGGAGACGGCTGGCGGTG	8082
QY	488	GATCCGGCCAACTCCACATCATCTACTTTGGTGTGCTGCTCAGAAACGGCCTCTGGAAAG	547
Db	8083	GACCCACACAGATGGGCGGGTGTGCTGTGGGCTCTGGCGC--GATGGCGGCGCTGTGGGCT	8139
QY	548	TCTACGACGGGGGGGGGTGACCTTTTCAAAGGCTGTGTTTCAGGGAACCTGGGACGTAC	607
Db	8140	AGCGACGATTCGGGGCGGCGCAGCTAGGGCGAAAGTGGCGTGTTCCTCCGACGCGCGCTGGCC	8199
QY	608	ATCCGAGACCCGAGTGTATCCAACGGGCTAACACAGCAACAGCAAGGACTCATGTGGTGT	667
Db	8200	GGTGCGACCGCAGCAATCATATTGTGGGCGGAGCAGCGGCTG--GGAGTCCCTTTGTGC	8256
QY	668	ACGTGCGACTCAACACGAGCAGACGCGGGGAGGCGACGCTCTGTAATCTTTTGGGACAG	727
Db	8257	GTTGTCAGACGACGCGAGTGGCAACCGGCAACCGCAACCGCGGATCTACTCTGGGCGGTG	8316
QY	728	GCTGATATCATCTACTGCTTCAAGTCTATGTGAGCAAGATGCCGGCTCCACGTAAGTGTCT	787
Db	8317	TC--CACCGAACAGACAGCGCTGTATGTGTCGAAGATGGCGGCGAGTTGGGCACCG	8373
QY	788	GTAACGGGGGCAACCGGAAATCTTTTCTCTACAAGCGGAAACTGACGCGACAGACAAG	847
Db	8374	GTTGGCGGGGCAACCGCGCGGCTGTGGGCC---GAGCGACATYGGCGCGGCGACGATG	8428
QY	848	G--CCTTTATCTGACTATTCCGATGGGACAGAGGCGCCTATGATGGACACTTGGCTAG	905
Db	8429	GCGCACTGGTACTGAGCTACGCGGACAGCCCGGGCCGAGCTGATGGCGGGGAGGCT	8488
QY	906	TGTGGAGGTACGACATTTCCAGGGGGAACCTGGAAAGACATCACCCCTGTCTGTGATCAG	965
Db	8489	TGTGGAATTTACGCGCGGACAGAGGGGCGCTGTGGCGGAGATTTAGCCCGATTTCGCGACGAG	8548
QY	966	ATCTATCTTTTGGCTTTGGCGGCTTTGGCCCTCGATTTTCAAAGGACGAGAACCCTGT	1023
Db	8549	CCAGTGGCGATGATTCGCTGGGCGGCGGTGGCGGTGATTCGGAACACCGCAGGT	8606

RESULT 9	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AE011809	AE011809	14520 bp DNA					
	<i>Xanthomonas axonopodis</i> pv. citri str. 306,	linear	BCT 29-MAY-2002				
	the complete genome.	section 187 of 469 of					
	AE011809 AE008923						
	AE011809.1 GI:21107967						
	<i>Xanthomonas axonopodis</i> pv. citri str. 306.						
	<i>Xanthomonas axonopodis</i> pv. citri str. 306.						
	Bacteria; Proteobacteria; gamma subdivision; <i>Xanthomonas</i> group;						

REFERENCE
AUTHORS

Xanthomonas.
1 (bases 1 to 14520)
da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J.,
Chambergro,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
Cursinho-Santos,J.R., El-Dorri,H., Ferreira,J.B., Ferreira,A.J.S.,
Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
Greggio,C.C., Gruber,A., Katuyama,A.M., Kishi,L.T., Leite
Jr.,R.P., Lemos,E.G.M., Lemos,R.S.I.N.M., Martins,E.C., Machado,M.A.,
Medeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Medeiros,J.,
Menck,C.E.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,V.R., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
Rossi,A.V., Sena,J.A.D., Silva,C., de Souza,R.F., Spíndola,L.A.F.,
Taktir,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kitaajima,J.P.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Comparison of the genomes of two Xanthomonas pathogens with
differing host specificities
Nature 417 (6887), 459-463 (2002).
22022145
12024217

2 (bases 1 to 14520)
da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J.,
Chambergro,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
Cursinho-Santos,J.R., El-Dorri,H., Ferreira,J.B., Ferreira,A.J.S.,
Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
Greggio,C.C., Gruber,A., Katuyama,A.M., Kishi,L.T., Leite
Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locati,E.C., Machado,M.A.,
Medeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Medeiros,J.,
Menck,C.E.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,V.R., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
Rossi,A.V., Sena,J.A.D., Silva,C., de Souza,R.F., Spíndola,L.A.F.,
Taktir,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kitaajima,J.P.

TITLE
JOURNAL

Direct Submission
Submitted (28-NOV-2001) Departamento de Biocinética, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
Brazil

FEATURES
Source

Location/Qualifiers
1..14520
/organism="Xanthomonas axonopodis pv. citri str. 306"
/strain="306"
/db_xref="taxon:190486"
/note="pathovar: citri"
87..2342
/gene="ceIa"
/note="XAC1770"
87..2342
/gene="ceIa"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="AA036634.1"
/db_xref="GI:21107968"
/translation="MIVIDPPTRRRSAAHWLTAVCLTLLVLFMTSEASABEPGP
DPMWHSVAIGGFFVGLVHPARBLAARDYGVGAADQAOAWLPTMLDGL
DMLNLGIDAPVADPADPNALYLAASTGTHERRAGNAYLRSPDCKQEPERDLPLCG
NOLGRANGERLAVDPHDGRVLLAGSTRGAGLRSDRGAHARVBDGPADLACATARN
HVSSEQVAGIAVFVDAASGAGATPRIVGVSTAGTSLYSVEDACRTVSAVAGOK
GLPSHAGNAGCWLYLSYGDREPDLMDAGALWNYATGQPREISIPQAPATGDF
GMGVAADPOHPVILASTRRPRPRPREVRSRGGGSAWPLAAAFDHSAPMTAH
ATPWICGLALIDPDSNHAFTVVGIGYASRNLOTSEGOPLWMDQDLLEVPILD
LISMAAHLLSLAGLDIDGRHDLPTNOQLYGRPRLTNESTIDGACQAQWVRSST
VDRKRNRIEALVSODGTHMAFASEPFGQAGTIAIADASQVAVWPDGQVHT
GDGRNRQORQVLEPTDAVVVADRDARWAAADRVSGRLKLESSGASVFPDQOVGS


```

gene      complement(3135. .4142)
CDS       /gene="MT1006.1"
          /complement(3135. .4142)
          /note="similar to GB:AL123456; identified by sequence
          similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="PE_PGRS family protein"
          /protein_id="AAK45254.1"
          /db_xref="GI:13880572"
          /translation="MSFVYVAPOLVSTAADARISATINTAATAATTQVLAAD
          EVSTAIALEGSHQHTVAISAQVAVQRFVLAISQASTYAVAAEAATPQLQIEQ
          ALGVINTPTPEALVGRKLIQDGAHAGPQTQAGAGGILMNGNGSGAPGQAGAG
          GAAGLIGNGAGAGTGGAGVSLARAGTAGAGRPVGGIGAGVGAGGAAGATTITH
          ASFNDHGVAVNPGSNVYVTFNGSGTVSVINPATNTVTGSPITIGNGPSGVAAPVYG
          LVEFTNFDSTNTSVIDPTNTVTGSPITVTGTAPTGVAVNPVIGEYVTVTFADPTVSVI
          S"
gene      4321. .4845
          /gene="MT1007"
          /note="identified by Glimmer2; putative"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAK45255.1"
          /db_xref="GI:13880573"
          /translation="MSAGETRADENSCVROHKSLLTGTORTPEAPCRPRVRLNLS
          ASRPLPMSPIVPOPHHRCNDRELTYTGPCNPPELGRAVFNRYPRHCRDRR
          AHLRAHTRKGRPTMAVPRKRSNTRSRSQMAAKTELGVTVAGAHKVPKRL
          LKAARGLIDFDR"
gene      complement(4864. .6294)
          /gene="MT1008"
          /complement(4864. .6294)
          /note="similar to GB:AL123456; identified by sequence
          similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="PE_PGRS family protein"
          /protein_id="AAK45256.1"
          /db_xref="GI:13880574"
          /translation="MGRPLNIVAVRISMSPVNAVAPOLVSTAADARISANTAN
          TAAATTOVLAADHDEVSTAIALEFSGHSHQVQVIAQVAAVOERVLALSQSSYVA
          VEAASATPQIQEALGVINTPTPEALVGRKLIQDGAHAGPQTQAGAGGILMNG
          GNGSGAPGQAGAGAGLIGNGAGTGGGGLPREAGANGAGGAGGMLRNGGAG
          GNGAGAGATTTCVAGGCGTGGCGHGLIGVCGHCGCGCGTGGCAVSLARAGT
          GAGGAGAGTGGAGVAGAGAGATTTHASFNDHGVAVNPGSNVYVTFNGSGTV
          SVTDPVNTVTGSDITDNGPSGVAAPVYGLVFTNFDSTNTSVIDPTNTVTGSPIT
          VGAGVAVNPGSNVYVTFNGSGTVSVINPATNTVTGSPITIGNGPSGVAAPVYG
          LVEFTNFDSTNTSVIDPTNTVTGSPITVTGTAPTGVAVNPVIGEYVTVTFADPTVSVI
          S"
gene      6614. .7306
          /gene="MT1009"
          /note="similar to PID:1465701 PID:1857994 GB:AL123456;
          identified by sequence similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="DNA-binding response regulator"
          /protein_id="AAK45257.1"
          /db_xref="GI:13880575"
          /translation="MSVRLVYVDDRAVRESLRSLSPNGYSVELAHDVEALDMVIA
          DRDVALVDVMPRLDGLVGRGLRGTGDDPLVLTARDVSERVAGADAGADVLP
          KPALELLARMAALIRTKRPAESMAARSDLLDPTVEVNRGQRISTRER
          ALLEMLIAMPRRVLRSLILEEYWGDFPTSGNALVEYVYGLRRTKTEADGEPLHTV
          KGVGLRTPP"
gene      7306. .8820
          /gene="MT1010"
          /note="similar to GP:4455704; identified by sequence
          similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="ABC transporter, ATP-binding protein"
          /protein_id="AAK45262.1"

```

```

          similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="sensor histidine kinase"
          /protein_id="AAK45258.1"
          /db_xref="GI:13880576"
          /translation="MMFRRDRAPLRATSSLSLRKRVALLMSVNAVYVLMFAVY
          AVISALYSIDINQLOSRAQLLIASGLAADPGKAIECTAASDVMAVNPQOSITYTA
          OPGQTLIPGAAGKAVIRGELEFMSRRTTADQVLAIRLNGSSLLISKSLKPTAEVAM
          KIRMWLLIVGGIGVAVAAVAGGVMPAGLRPGLTEAERVAARDLRLPTVFGSDE
          LARTRAPNMLRLALESREKARLVTDGHELRPTLSLRRTNVELMASMAPCARL
          PROENVDLRADVLAQIEELSTLVGLDVLDSKQDAGEVHEPYDMAVDVRSLEKRRR
          RNDILFEDVEITVMQVGYDPAGLSRMALNMDAAKSPGPGVGRVLSQDASHLEY
          VSDRGQIVQERLVEFEREYSSASARALPGSGGLAIYKQVVLNMGGLRTEEDTPG
          GPPGGSIVYLLPGRMMPQLPGATAGARSTDIENRSANVISVESQSTRAT"
gene      8918. .10258
          /gene="MT1011"
          /note="similar to GP:4455705; identified by sequence
          similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="heat shock protein HtrA, putative"
          /protein_id="AAK45259.1"
          /db_xref="GI:13880577"
          /translation="MTNHRYSPPPOQPGTPEYAOGQOQTVSQQFDMRYPPSPPOPT
          OYROPYEAALGTRPGLIPGVITPMPPGMVORPAGALAIQAVTIAVNSGIGGA
          ASVYCPNRPADPSCGPVAAASAPSRANMPGVSVEQVAAKVPVSVLLFEDLGRS
          BEGSGITILABEGLITLNNHYIAAAKRPGLSPPTKTYVFSGRTAPPTVVGADPPSD
          IAVNRQVGSVLPISLSSSDLRQVQGLVMAQLGSPLEGITVTTGTVSLADSPVSGT
          EAGNOMTVDAIQTDAAINPNSGGLVMMNAQLGVNSAIAITLADSDAOSGSLGL
          GFAIPVDOAKRIADELITSTKASHASISLGVQVNDKDTPAKTIQEVVAGAAANAGVPK
          GVYVTVKDDRPINSADALVAARSKAPGATVALTRQDPGSGSRVQVTLGRKEQ"
gene      10255. .10803
          /gene="MT1012"
          /note="similar to GB:AE000511 PID:2313931; identified by
          sequence similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="molybdopterlin biosynthesis protein"
          /protein_id="AAK45260.1"
          /db_xref="GI:13880578"
          /translation="MKKVAOCSKLGTVAPMEORAEVYGRALVVVDPRPHGDE
          HSGPLVTLELTFAGFVNDGVAVASDEVIRNAINTAVIGVDIVSVSGCTGVTPRDY
          TPBTRDITLDEFLIGAEIRASGLSAGTIVDAGLSKLAGVSGSTLVNLAASRAVR
          DKMATLNPALAQITGLSSLET"
gene      complement(10823. .11278)
          /gene="MT1013"
          /complement(10823. .11278)
          /note="similar to GB:L29458 SP:P23867 GB:U08371 GB:K52114
          PID:43136; identified by sequence similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="large conductance mechanosensitive channel"
          /protein_id="AAK45261.1"
          /db_xref="GI:13880579"
          /translation="MLKGRKFLRANGNIYDLAVAVVIGTAFTALYTKPDSITPLIN
          RIGVNAOSQVGLIRIGITGGQRTIDLNVLSAIAINFLLAFVAVFLVLYNLRKKE
          VEOPGTVQVLLTEIRIDLQNTGDSPPGRHGGRTGTPSPGPRASIESQ"
gene      11601. .12347
          /gene="MT1014"
          /note="similar to GB:AE000782; identified by sequence
          similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="ABC transporter, ATP-binding protein"
          /protein_id="AAK45262.1"

```



```

/db_xref="GI:13880580"
/translation="NMROPIVOLSNLSMTFREGERTROYLDITFDEPEFVALIGO
SSGSKTILNLSGIEKPTPTGVTNGITFOTKTRERDLFRPDIGIYFQEPNLIPT
LTVLENTITPOLIACVSORKAAVARDLLEKGMADRETFPDKLSGGQONALISRA
LAHNPMLVLADEPTGNLSDTDKVDVLDLITROAGKTLIMATHSPSMTQADRVVN
LOGGRILIPAVNENQTDOPASTILPLTSE"
gene
12340..14907
/gene="MT1015"
12340..14907
/cds
12340..14907
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11

Query Match      3.4%  Score 92.2; DB 1: Length 16168;
Best Local Similarity 43.9%; Pred. No. 2.9e-05;
Matches 443; Conservative 0; Mismatches 563; Indels 3; Gaps 1;

OY 1370 GAGCGCATGAGGAATYTCCTCCAGAGACCTTGCTGACCCGCGGAGAGCTA 1429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1653 GAAGTGCCGCTCCGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1430 TTGGCCGCACTGCGAGACGACGAGCTTACCTTCCAGCAAGAACCTCGGACA 1489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1593 CCGCGCGCATTCAGACCGCTGTGCGCTGCGCGCTGCGCGCGCGCGCGCGCAC 1534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1490 TCGCCGACAGCGTCTGCGACGCGCCACATGGGCGCACCTGACGAGCGTCACTACGCC 1549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1533 GCCTTGCGCGCGCGCTCCCGCGCGCTTCGATGCGACGCTTGCGCGCGCGCGCG 1474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1550 GGGACTCGGTCAAGACCGTGTCCGCGTGGCAACACCGCGCGCGCGCGCGCG 1609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1473 GTAGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1610 ATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1413 GCGCCCTTCGCGCGCTGCCAACACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1670 GCGCGACGCGTGGCTATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1353 ATCACTTCCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1730 GCGGTGACGCGTCCGAGTTCAGGGGAGCTTTCCTTCGAGCGTCCGCGCG 1789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1293 GGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1790 GCGGTATGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1849
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1233 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1850 TTTTACGTCAGCAAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1173 GCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1910 GGGACGATCGGGATTCGCTGTACCGCGCGCGCGCGCGCGCGCGCGCGCG 1969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1113 GATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1970 GAGTGGCGCATTTCCGCTCCACAGACTGGGAGACGCTTGGCCAGTCTCC--ACC 2026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1053 GACCGTGATGCTCCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2027 GCGCTGACCAACGATACGATCGCTTGGTGGGCTCAGGCTCGAAGCTGGAACCTG 2086
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1993 GCGCGTACGCGCATTCGCTTTCCTCCCGTGGCGAACAAGCGCGCGCGCGCG 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2087 TATGCTTGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1933 GACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2147 TCGTGGAGGACATCCAGGGCTCCAGGGGCTTGGGCTCCATCCAGACACCAAGTCCG 2206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
873 ACCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

OY 2207 GCGAGCGGACGACCGCGCGCGCAAGTCTAGTGGGACCAAGCGCGCGCTTTAC 2266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 813 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2267 GCTCAGGAACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 753 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2327 AGCAGTACCTCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 693 AGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
AL161725
LOCUS
DEFINITION
AL161725 181179 bp DNA linear PRI 16-JAN-2001
Human DNA sequence from clone RP11-165F24 on chromosome 9. Contains
the 3' end of the gene for a novel protein (similar to Drosophila
CG6630 and CG11376, KIAA1058, rat TRG), an RPL12 (60S ribosomal
protein L12) pseudogene, ESTs, STSs, GSSs and a CpG island,
complete sequence.
AL161725
VERSION
AL161725.13 GI:10045359
KEYWORDS
HTG; CpG island; KIAA1058; RPL12; TRG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 181179)
REFERENCE
Laird,G.
Direct Submission
Submitted (11-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 9, 2000 this sequence version replaced gi:9864498.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information
on the WormPeP database can be found at
http://www.sanger.ac.uk/projects/Celegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
RP11-165F24 is from the library RPL11-11 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-165F24 The true
right end of clone RP11-5906 is at 17684 in this sequence. The true
left end of clone RP11-31F19 is at 158244 in this sequence.

FEATURES
source
1..181179
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-165F24"
/clone.lib="RPL11-11"
misc_feature
9..757
/feature="match: GSS: Em:AQ417012"
misc_feature
22..474

```



```
/note="match: GSS: Em:AQ318836"
56..480
/note="match: GSS: Em:AQ190327"
809..1072
/note="MIR repeat: matches 8..262 of consensus"
1121..1289
/note="L2 repeat: matches 2590..2750 of consensus"
1881..9918
/gene="ba165f24.1"
join<1881..1998,4093..4163,5291..5429,6048..6149,
10073..10168,10840..11074,13634..13795,16376..16548,
20194..20289,24334..24429,30648..30797,33009..33122,
38781..38936,40793..40932,48645..48814,51931..52070,
54264..54446,54812..54941,55911..55998,60748..60844,
62225..62359,63565..63717,66029..66187,67738..67838,
68646..68838,73108..73251,75149..75280,75738..75872,
77290..77379,80233..80469,83647..83790,85874..85980,
97380..97550,98022..99118)
/gene="ba165f24.1"
/product="ba165f24.1.1 (novel protein similar to
Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
1))"
/note="match: CDMAS: Em:AB037816
match: ESTs: Em:AW504701 Em:AW50006 Em:AW500946 Em:HB9666
Em:W30966 Em:AW49621 Em:A1524110 Em:AJ394589 Em:AA146391
Em:AW61668 Em:AA798223 Em:AW504408 Em:AW337875
Em:A1927976 Em:AW500913 Em:AA374756 Em:T63639 Em:AA321435
Em:AA402794 Em:AA212554 Em:AA146403 Em:AA365573 Em:N6754
Em:CO1477 Em:CO1637 Em:T80502 Em:AW411026 Em:A1192903
Em:AA170814 Em:A1280154 Em:AA251655 Em:AA007697
Em:A1660115 Em:AA893933 Em:A1422073 Em:AA552743 Em:N48396
Em:AA401587 Em:N30451 Em:N57465 Em:AA401548 Em:A1862755
Em:AA400038 Em:AW511478 Em:AA207838 Em:A1990239
Em:AA659397 Em:A1693055 Em:AA213632 Em:R70724 Em:A158119
Em:N30504 Em:A1480327 Em:N30464 Em:A1769537 Em:AA742950
Em:A1621032 Em:AA251553"
/evidence="not_experimental"
join<1881..1998,4093..4163,5291..5429,6048..6149,
10073..10168,10840..11074,13634..13795,16376..16548,
20194..20289,24334..24429,30648..30797,33009..33122,
38781..38936,40793..40932,48645..48814,51931..52070,
54264..54446,54812..54941,55911..55998,60748..60844,
62225..62359,63565..63717,66029..66187,67738..67838,
68646..68838,73108..73251,75149..75280,75738..75872,
77290..77379,80233..80469,83647..83790,85874..85980,
97380..97550,98936..99118)
/gene="ba165f24.1"
/product="ba165f24.1.2 (novel protein similar to
Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
2))"
/note="match: ESTs: Em:AW337875 Em:CO1477"
/evidence="not_experimental"
join<1881..1998,4093..4163,5291..5429,6048..6149,
10073..10168,10840..11074,13634..13795,16376..16548,
20194..20289,24334..24429,30648..30797,33009..33122,
38781..38936,40793..40932,48645..48814,51931..52070,
54264..54446,54812..54941,55911..55998,60748..60844,
62225..62359,63565..63717,66029..66187,67738..67838,
68646..68838,73108..73251,75149..75280,75738..75872,
77290..77379,80233..80469,83647..83790,85874..85980,
97380..97550,98936..99002)
/gene="ba165f24.1"
/codon_start=2
/evidence="not_experimental"
/product="ba165f24.1.2 (novel protein similar to
Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
2))"
/protein_id="CAC22149.1"
/db_xref="GI:12274836"
/translation="NLLYYPQRLNFVNKLASARNITTIQWCGEDASNAMPYFGK
SSGPEFLQEVYAVTYHNKSPDYEEVKIKLPKLVNHHLLTFYHISCOQKASV
EPFLGSMILINERLOGSCYCLPALEKIPNPSMHAEEVPLQNPPEIKAECHG
VFNIEQVAVSSVHTQDNHLEKFFTTLCSHLESQVTFPIRVLDQKISMALEHLEKLSII
```

```
CDS
10073..10168,10840..11074,13634..13795,16376..16548,
20194..20289,24334..24429,30648..30797,33009..33122,
38781..38936,40793..40932,48645..48814,51931..52070,
54264..54446,54812..54941,55911..55998,60748..60844,
62225..62359,63565..63717,66029..66187,67738..67838,
68646..68838,73108..73251,75149..75280,75738..75872,
77290..77379,80233..80469,83647..83790,85874..85980,
97380..97550,98022..99118)
/gene="ba165f24.1"
/note="match: proteins: Tr:Q9VP19 Tr:Q20487 Tr:Q9PU04
Tr:Q9VM04 Tr:Q63603"
/codon_start=2
/evidence="not_experimental"
/product="ba165f24.1.1 (novel protein similar to
Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
1))"
/protein_id="CAC22148.1"
/db_xref="GI:12274835"
/translation="NLLYYPQRLNFVNKLASARNITTIQWCGEDASNAMPYFGK
SSGPEFLQEVYAVTYHNKSPDYEEVKIKLPKLVNHHLLTFYHISCOQKASV
EPFLGSMILINERLOGSCYCLPALEKIPNPSMHAEEVPLQNPPEIKAECHG
VFNIEQVAVSSVHTQDNHLEKFFTTLCSHLESQVTFPIRVLDQKISMALEHLEKLSII
CLNSSLREPLVLEHLVLDKLPOLSVQPVNIAGOTANFSQAFESVVAIANSLHNSKD
LSKDQGRNCLASVYHYVRLPEVORDVPCAPALADPRRSYHYHGRFSAASVSKL
LQARVSSSNPLDAGHSADEEVKNINSK1ADNRCSMSYCGSSSDAPSPAPR
PASKKHEHEHLALQWVSTGWRREYFKAMFFELLYKSMQHNNMKRSPFRT
FSDRFMDITTTIVNVTSETIALVAPKQENQAEKMN1SLFPLDLSLDKRGV
NLIIRHYSQLSAKSLNPLTILSMRLEFLILCSHEHYLN1LNFNNADLPDPS
SSONSSCSSPDDOKIASMFDTSEVROOHPFTGLTFLEIALAALDAEGEGISVKVKA
VSAIHSLSSHDLPKPCVKEVVKIAALYPLVGLITDIALPOLDFYADRRYTS
GSDEREGALINONVALA1AGNNFLKTSGLVSLSPKQVNMADTRNLMLCFL
WIMKNADOSTLITKMTADLPSTOLNRTDLDFLCVFEKKQSSDDKSTOVLQKSD
VVARLEBALREGAGEMRRARRRNDPFLNTRKKTQTHIRONENLDTKA
E1DQELISGNLATEHNLITLMOENITDASSALDCKDLSGLVRLVANS1LNCQST
TYLTHGATLRLALIAFGLDEEVEQFDLCHOVRLSSSMQVTRQACATLYL
MFESECATNFARVKQVMTASLISVGRAPDNEEHLRS1LTIILASEEDPAMQPT
PFOVEYLICNLSNIYDVKKREDEDEMDLMDYRTRAKYQASPOLR1TWLQMA
EKHTKKCYTEAMCLVNAALVAE1LSLDEHYS1PVSVEFQNS1SVLEESVE
DPLSDPDCVQAGVTFEYK1KIDTPVDKTLDPKAV1QITVEEYFDEYEMKDRV
E1SHRLEAFVGOFGAEFEV1KIDTPVDKTLDPKAV1QITVEEYFDEYEMKDRV
TYEKENFLRREMYTTPLEGRPCELEHORYRNTVLTTHMAFYIKTRISV1QKE
FULTPLEVLA1RHMKKTQLAVALINQEPDAMQNLQSGVATVNOGLEVACVFL
AE1PADPKLYRHHNKLRLCFEKEIFMCGAVENKRNRLITADOREYQOELKKVYNLKE
NLRPMTERK1RPLLYKPIRVESSOKRTICLSAVYQLOINRPFDE"
join<1881..1998,4093..4163,5291..5429,6048..6149,
10073..10168,10840..11074,13634..13795,16376..16548,
20194..20289,24334..24429,30648..30797,33009..33122,
38781..38936,40793..40932,48645..48814,51931..52070,
54264..54446,54812..54941,55911..55998,60748..60844,
62225..62359,63565..63717,66029..66187,67738..67838,
68646..68838,73108..73251,75149..75280,75738..75872,
77290..77379,80233..80469,83647..83790,85874..85980,
97380..97550,98022..99082)
/gene="ba165f24.1"
/note="match: proteins: Tr:Q9VP19 Tr:Q20487 Tr:Q9PU04
Tr:Q9VM04 Tr:Q63603"
/codon_start=2
/evidence="not_experimental"
/product="ba165f24.1.1 (novel protein similar to
Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
1))"
/protein_id="CAC22148.1"
/db_xref="GI:12274835"
/translation="NLLYYPQRLNFVNKLASARNITTIQWCGEDASNAMPYFGK
SSGPEFLQEVYAVTYHNKSPDYEEVKIKLPKLVNHHLLTFYHISCOQKASV
EPFLGSMILINERLOGSCYCLPALEKIPNPSMHAEEVPLQNPPEIKAECHG
VFNIEQVAVSSVHTQDNHLEKFFTTLCSHLESQVTFPIRVLDQKISMALEHLEKLSII
CLNSSLREPLVLEHLVLDKLPOLSVQPVNIAGOTANFSQAFESVVAIANSLHNSKD
LSKDQGRNCLASVYHYVRLPEVORDVPCAPALADPRRSYHYHGRFSAASVSKL
LQARVSSSNPLDAGHSADEEVKNINSK1ADNRCSMSYCGSSSDAPSPAPR
PASKKHEHEHLALQWVSTGWRREYFKAMFFELLYKSMQHNNMKRSPFRT
FSDRFMDITTTIVNVTSETIALVAPKQENQAEKMN1SLFPLDLSLDKRGV
NLIIRHYSQLSAKSLNPLTILSMRLEFLILCSHEHYLN1LNFNNADLPDPS
SSONSSCSSPDDOKIASMFDTSEVROOHPFTGLTFLEIALAALDAEGEGISVKVKA
VSAIHSLSSHDLPKPCVKEVVKIAALYPLVGLITDIALPOLDFYADRRYTS
GSDEREGALINONVALA1AGNNFLKTSGLVSLSPKQVNMADTRNLMLCFL
WIMKNADOSTLITKMTADLPSTOLNRTDLDFLCVFEKKQSSDDKSTOVLQKSD
VVARLEBALREGAGEMRRARRRNDPFLNTRKKTQTHIRONENLDTKA
E1DQELISGNLATEHNLITLMOENITDASSALDCKDLSGLVRLVANS1LNCQST
TYLTHGATLRLALIAFGLDEEVEQFDLCHOVRLSSSMQVTRQACATLYL
MFESECATNFARVKQVMTASLISVGRAPDNEEHLRS1LTIILASEEDPAMQPT
PFOVEYLICNLSNIYDVKKREDEDEMDLMDYRTRAKYQASPOLR1TWLQMA
EKHTKKCYTEAMCLVNAALVAE1LSLDEHYS1PVSVEFQNS1SVLEESVE
DPLSDPDCVQAGVTFEYK1KIDTPVDKTLDPKAV1QITVEEYFDEYEMKDRV
E1SHRLEAFVGOFGAEFEV1KIDTPVDKTLDPKAV1QITVEEYFDEYEMKDRV
TYEKENFLRREMYTTPLEGRPCELEHORYRNTVLTTHMAFYIKTRISV1QKE
FULTPLEVLA1RHMKKTQLAVALINQEPDAMQNLQSGVATVNOGLEVACVFL
AE1PADPKLYRHHNKLRLCFEKEIFMCGAVENKRNRLITADOREYQOELKKVYNLKE
NLRPMTERK1RPLLYKPIRVESSOKRTICLSAVYQLOINRPFDE"
complement(2149..2653)
```

```
misc_feature
Query Match 3 4%: Score 92: DB 9: Length 181179:
Best Local Similarity 44.3%: Pred. No. 2.4e-05:
```

Matches 415; Conservative 0; Mismatches 520; Indels 1; Gaps 1;

OY	1450	CAGGGCTTCAACCTTTGGGACAGAGAAAGACCTCGGGGACATGCGCGGACAGAGGCTGTGGG	1509
Db	33943	CACCAACCTCCACACATACACACACACCACTCCACACCAACAGACATCTTACCATACACAC	34002
OY	1510	AACGCCACATAGGGGACCACTCGAGAGAGGTGATACACCCGGGAATCGGTTCAAGAGGT	1569
Db	34003	CACCTTCACACATACACACACACCACTCCACACATACACACTCTTTCACCATACCAT	34062
OY	1570	CGTCCGCTCGGCAACACCCGCGGACGACAGGTGGCCATCTGTCCGAGGGGGGGC	1629
Db	34063	CACCAACCACTTCATACATACACCACTCCACATCCACATCCACCACTTCACCATCACAC	34122
OY	1630	GACCTGGAGCATTCGACTACGGGGCCGACACGTCATGAAAGGGGACAGGTGGCTATTTC	1689
Db	34123	CACCTCCACACATCACACCACTCTCTTCAACCAACCAACATCTTTCACCGTACCAACAC	34182
OY	1690	GCCCGACGGGACACGATCTCTGTGTGACAGCGCTCTGTCGGGCGTGACAGCGCTCGAGTT	1749
Db	34183	CTCCACACACACACACCTCTCTCCACACCAACCACTCTCTACATACACACTCTTTCAC	34242
OY	1750	CCAGGGCAGCTTTGCTCCGTCTCGAGCCTGCCGCGGGCGCGTACCTCGCTGGACAA	1809
Db	34243	CATCCACATACACACACCTCTCCACCATCACACCAACCACTCCACATACCACTCACCTTC	34302
OY	1810	GAAGGCAACAGGCTCTCTACGGCGGGCTCC-GGATCGACCTTTACGTACACAAAGACA	1868
Db	34303	CACCAACCAACGATCTTTACATACCAACCAACCACTCTCTTACACATCACACCACTCT	34362
OY	1869	CCGGCAGCAGCTTTCACGGCGGGGCCAAGCTTGGGACGGGACAGGACGATCCGGGATATCG	1928
Db	34363	CCACCATACCAACCAACCACTCTCCACATACCAACCACTCTCCACCAACCAACGATCTTCA	34422
OY	1929	CTGCTACCCGACACACCGGGGACAGCTTGATGTCTCGACGAGCTCGGCAATATTCGGT	1988
Db	34423	CCATCACCACGCACTCTCACATACACCACTCTCCACATCACCAACCACTCTCCA	34482
OY	1989	CCACAGACTCGGGGACGAGCTTTGGGCAAGTCTCACGGCCCTTGACCAACACTCTACAGA	2048
Db	34483	CCACCATCACCAACCACTCTCACATACCACTCTCCACCAACCAACGATCTTTCACCA	34542
OY	2049	TCCGCCCTGGGTGTGGCTCAGGCTGAACTGGAACTGTAATCTTGGGACACGGGCGCT	2108
Db	34543	TCACCAACCACTCCACCAACCAACCACTCTCACATACCACTCTCACCAACCACTCTCACCA	34602
OY	2109	CAGGGGCTCGCTCTACGCCAAGTGAAGAGCGGGCCTCTTGAAGGACATCCAGGCT	2168
Db	34603	CCAGCATCTTTCACATACACCACTCTCCACCACTCACCACTCCACATCACCACTCA	34662
OY	2169	CCCAGGGCTTGGGCTTCATCGACACACCAAGTGGCGGGGACGAGGCGACACCGCGGGC	2228
Db	34663	CTTCCACCATCACCAACCACTCTCACATCACCACTCTCCACCACTCTCCACATCACCACT	34722
OY	2229	AAGTTACGTGGGACCAAGCGGCGGGGCTTTTACGCTCAGGGAACCGTCGGCGGG	2288
Db	34723	CCTCCACCATCACCACTCTCACACCACTCTCACCACTCTCCACCACTCTCTTCA	34782
OY	2289	GCACGGGCGGGACTTCTCTGTGACCAAGACAGAGCAGACGATCTTTCGCGCAGCT	2348
Db	34783	CCATCACCAACCACTCACCACTCTCACCACTCTCACCACTCTCCACCACTCTCCACATCA	34842
OY	2349	CGAGCACCACGCTGAGGTGTGAGGCTTGTATTCACGA	2384
Db	34843	CCACCTCTCCACCACTCACCACTCTTACCTTACCTCACCA	34878

Search completed: May 4, 2003, 11:09:56
Job time : 5743.01 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 04:10:24 ; Search time 374.329 Seconds
(without alignments)
16303.594 Million cell updates/sec

Title: US-10-026-994-1
Perfect score: 2710
Sequence: 1 ccacgcgtccgcagcagtc.....ctattgaaaaaaaaaaaaa 2710

Scoring table: IDENTITY NUC
Gapop 10.0, Gapept 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	600.4	22.2	1103 21 AAF14988	Trichoderma reesei
2	326.6	12.1	406 21 AAF15118	Trichoderma reesei
3	88	3.2	4403765 22 AAI19983	Mycobacterium tuberculosis
4	87.2	3.2	65140 22 AAD17184	Streptomyces novus
5	87.2	3.2	125401 22 AAD17186	Streptomyces novus
6	78.4	2.9	1833 15 AAO64206	snbr gene encoding
7	71.4	2.6	4403765 22 AAI19983	Mycobacterium tuberculosis
8	70	2.6	2744 16 AAO98470	MSP1-containing P
9	69.8	2.6	3015 20 AAZ32027	Human METH1 relate

10	69.8	2.6	3015 22 AAC90084	246970 cDNA clone.
11	69.4	2.6	4411529 22 AAI19988	Mycobacterium tuberculosis
12	66.8	2.5	1107 22 AAF74540	Nucleotide sequence
13	66.8	2.5	2712 20 AAZ06825	Streptomyces albid
14	66.8	2.5	2712 22 AAF74538	Nucleotide sequence
15	65.4	2.4	1140 19 AAV41733	Codon-optimised RA
16	64.2	2.4	1635 22 ABA49946	Human breast cell
17	64.2	2.4	1635 22 ABA67865	Human foetal liver
18	64.2	2.4	1635 22 ABA34921	Probe #13387 for g
19	64.2	2.4	1635 22 AAK16270	Human brain expres
20	64.2	2.4	1635 22 AAK42016	Human bone marrow
21	64.2	2.4	1635 22 AAI22780	Probe #12713 for g
22	64.2	2.4	1635 22 AAI48082	Probe #16768 used
23	64.2	2.4	1635 22 AAI08454	Probe #8445 used t
24	64.2	2.4	1635 22 ABA16047	Human genome-deriv
25	64.2	2.4	1973 22 ABA44805	Human breast cell
26	64.2	2.4	1973 22 ABA55261	Human foetal liver
27	64.2	2.4	1973 22 ABA25005	Probe #3471 for ge
28	64.2	2.4	1973 22 AAK03514	Human brain expres
29	64.2	2.4	1973 22 AAK28970	Human bone marrow
30	64.2	2.4	1973 22 AAI13556	Probe #3604 used t
31	64.2	2.4	1973 22 AAI34918	Probe #3489 for ge
32	64.2	2.4	1973 22 AAI03446	Probe #3437 used t
33	64.2	2.4	1973 22 ABA03504	Human genome-deriv
34	64.2	2.4	114955 20 AAF25795	Human adenosine A1
35	63.6	2.3	3849 22 AAF22159	S. chromomallus ac
36	63.6	2.3	1833 23 AAN80309	Drosophila melanog
37	62.8	2.3	1596 17 AAX22316	Nocardiopsis sp. p
38	62.8	2.3	4451 23 ABA92722	Trichoderma reesei
39	62.8	2.3	4451 24 AAT172045	Chrysosporium CBH1
40	62	2.3	1806 23 AAS51433	Pseudomonas aerugi
41	60.6	2.2	2291 9 AAN80309	Entire amylase gen
42	60	2.2	4411529 22 AAI19988	Mycobacterium tuberculosis
43	59.8	2.2	3946 18 AAT93610	Myobacterium tube
44	59.6	2.2	2271 11 AAO06844	Amlyase gene from
45	59.6	2.2	2367 19 AAV64557	M. tuberculosis Im

ALIGNMENTS

RESULT 1	AAFI4988	standard; cDNA: 1103 BP.
ID	AAFI4988	
AC	AAFI4988;	
DT	13-MAR-2001	(first entry)
DE	Trichoderma reesei	EST SEQ ID NO: 7511.
KW	Multiple gene expression; filamentous fungal cell; EST;	
KW	expressed sequence tag; Fusarium venenatum; Aspergillus niger;	
KW	Aspergillus oryzae; Trichoderma reesei; identification; recombination;	
KW	culture condition; environmental stress; spore morphogenesis;	
KW	metabolic pathway engineering; catalytic pathway engineering; ss.	
OS	Trichoderma reesei.	
PN	WO200056762-A2.	
PD	28-SEP-2000.	
PF	22-MAR-2000; 2000WO-US07781.	
PR	22-MAR-1999; 99US-0273623.	
PA	(NOVO) NOVO NORDISK BIOTECH INC.	
PA	(NOVO) NOVO NORDISK AS.	
PI	Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;	
DR	WPI: 2000-594572/56.	

Monitoring differential expression of genes in filamentous fungal cells uses fluorescence-labeled nucleic acids isolated from the cells and a substrate of expressed sequence tags -

Claim 89; Page 3034; 3161pp; English.

The present invention describes a method for monitoring differential expression of genes in a first filamentous fungal (FF) cell relative to expression of the same genes in one or more second filamentous fungal cells. The method uses fluorescence-labeled nucleic acids isolated from the FF cells and a substrate of expressed sequence tags (EST). The ESTs are used in the methods for monitoring differential expression of genes in a first filamentous fungal(FF) cell relative to expression of the same genes in one or more second filamentous fungal cells. Monitoring the global expression of genes from FF cells allows the production potential of the microorganisms to be improved. New genes may be discovered, possible functions of unknown open reading frames can be identified and gene copy number variation and stability can be monitored. The expression of genes can be used to study how FF cells adapt to changes in culture conditions, environmental stress, spore morphogenesis, recombination, metabolic or catabolic pathway engineering. Using ESTs provides several advantages over genomic or random cDNA clones including elimination of redundancy as one spot on an array equals one gene or open reading frame, and organisation of the microarrays based on function of the gene products to facilitate analysis of the results. AAF07478 to AAF11247 represents ESTs from *Fusarium venenatum*. AAF11248 to AAF11854 represents ESTs from *Aspergillus niger*. AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*, and AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are all specifically claimed in the present invention.

Sequence 1103 BP; 230 A; 322 C; 307 G; 223 T; 21 other;

Query Match	22.2%	Score 600.4	DB 21	Length 1103
Best Local Similarity	85.8%	Pred.No. 4.8e-122		
Matches 810; Conservative	0	Mismatches 96	Indels 38	Gaps 12

Oy	984	GGGGCTTTGGCGTCGATTTTGCAGAAAGCCGAGAAACCTTTGTCCTTTTGAACCTTT	1043
Db	1	GGGGCTTTGGCGTCGATTTTGCAGAAAGCCGAGAAACCTTTGTCCTTTTGAACCTTT	60
Oy	1044	GGTGGCCAGATGCTCAGCTGTGTTGGTGCAGCCGACTTGGGCAACATGAGCCCGATCT	1103
Db	61	GGTGGCCAGATGCTCAGCTGTGTTGGTGCAGCCGACTTGGGCAACATGAGCCCGATCT	120
Oy	1104	GGGGGTGGGGAGGTATCCGACTGAGACCAATTACTACAGCATCTCACTCCCAAGAC	1163
Db	121	GGGGGTGGGGAGGTATCCGACTGAGACCAATTACTACAGCATCTCACTCCCAAGAC	180
Oy	1164	CGTGGATCAAGAACACACTTTATGATGTGACGAGCGAGTCACCGTCGATGGTCTCATCA	1233
Db	181	CGTGGATCAAGAACACACTTTATGATGTGACGAGCGAGTCACCGTCGATGGTCTCATCA	240
Oy	1224	AGCGCCCTCGCGCTGGATGATGAGTCTCTCTGGAGATTGACCCAGCAGACCAACCTGGC	1283
Db	241	AGCGCCCTCGCGCTGGATGATGAGTCTCTCTGGAGATTGACCCAGCAGACCAANNNCTGGC	300
Oy	1284	-TTCACGGCACCCGGAATGACAATCTTTGGCGGCACAGATCTCACCACATGGGAGCACGGC	1342
Db	301	TTTTACGGCACCCGGAATGACAATNTTTGGGGGCGACGATTCACCACATCGGGACACGGC	360
Oy	1343	C--ACATGTGTCAATTCATTCAC-TGGCAGAGCGCAT---CGAGGAATTCCTCGGTCCA	1395
Db	361	CCACAAATGTGTCAATTCATTCATTCCTGGCAGAGCGGATTCGAAGAAATTTTCCGTCTAA	420
Oy	1396	GGACCTTGGCTTGCACCC-GCGGGAAGCGAGCTATTGGCCGCA--GTTCGAGAGCA-CA	1451
Db	421	GGACCTTGGCTTTCACCCGGGGGGAAGGAGCTTTTGGCCGCAATGCCGAGAGCAGNCA	480
Oy	1452	ACGGCTTCACCTTT--GCCAGCAAGAACGACTCGGGACATCGCCGACAGAGGCTCGGC	1509
Db	481	ACGGCTTTACCTTTTGGCCAGCAAGAACGACTTTGGGACATTGGCCGACAGAGCGTTGGGC	540

Qy	1510	AACCCCAACATGGGGCCACCTCCGACG - AGCGTGACGTACGCGGGGAACATCGGTCAAGAGNC -	1567
Db	541	AACCTCCACATGGGACCACTTCGACGAAGGCTGACGTACCGCGGGAACCTCGGTCAAGAGCC	600
Qy	1568	- GTCGTCCGCGGTGGGCAACAC - GCCGGACGCAACAGGTGGCCATCTCGTCCGACGGG	1625
Db	601	GTTCGTCCGCGGTGGGCAACACCGCGCGGACGCAACAGGTGGCCATTTTTCGAAAGC	660
Qy	1626	GCGCGACGTGGAGCATCGACTACGCGGCGACACGTCACT - - - - -	1655
Db	661	GCGCGGCCCCGACGCTGGGAACCAATTTGAACTACGCTGGGNTCCGAACACGTTTTCATTGG	720
Qy	1666	GAACGGGGGACGCTGGGCGCTATTGGGCGGACGCGACACAGATCTCTGTGTGACCGCTC	1725
Db	721	AAACGGGGGGCGGCTGGCGCTATTGGGCGGACGCGACACAGATCTCTGTGTGACCGCTC	780
Qy	1726	GTCGCGCGTGCACGCGCTCGGAGTTCGACGGGACAGCTTTGGCTCGCTCGGTGAGGCTGCCGC	1785
Db	781	GTCGCGCGGTGCAGCGCTCGGAGTTCGACGGGACAGCTTTGGCTCGCTCGGTGAGGCTGCCGC	840
Qy	1786	GGGCGCGCTATCGCTCGGACAGAAGACCAACAGCGTTCATCGCCCGGCTCCGATC	1845
Db	841	GGGCGCGCTATCGCTCGGACAGAAGACCAACAGCGTTCATCGCCCGGCTCCGATC	900
Qy	1846	GACCTTTTATAGTCAGCAAGAGACACCGGACAGACGTTTCACGCGCG	1889
Db	901	GACCTTTTATAGTCAGCAAGAGACACCGGACAGACGTTTTCACGCGCG	944

RESULT 2

ID AAF15118 standard; CDNA; 406 BP.

AC AAF15118;

DT 13-MAR-2001 (first entry)

DE Trichoderma reesei EST SEQ ID NO: 7641.

KW Multiple gene expression; filamentous fungal cell; ESF;
KW Multiple sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.

OS Trichoderma reesei

PN: W0200056762-A2.

PD 28-SEP-2000.

PF 22-MAR-2000; 2000WO-US07781.

PR 22-MAR-1999; 99US-0273623.

PA (NOVO) NOVO NORDISK BIOTECH INC.

XX

XX

XX

PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -

PS Claim 89; Page 3084-3085; 3161pp; English.

CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from


```
Db 3936636 GCTGCGCGCGCGCTCCGGTGCCTGCGGACCGCCTTTGGCCCGCGCTGCGAGCGGTGCGG 3936577
OY 1971 ACGTGGGCAATATTCGGCTCCACAGACTCGGGACAGACCTTTGGCCAAAGTCTCCACCGCCC 2030
Db 3936576 TTAGCGCCCGCCACCGCCCGCCGCGCCGCGCCCGCCACCGGACACACCGCTGCGCCCT 3936517
OY 2031 TGACCAACACCTACAGATGCGCCCTGGGTGGGGCTGAGCTCGAATCGAATCGAATGATG 2090
Db 3936516 TGACCTCTCTTGGCCCGCGCTGCGGACCTGTGTGGGGCCGCTGCGCGCGCGCGCG 3936457
OY 2091 CTTTGGGACCGCGCCCTTCAGAGGGGCTGCGCTCTTCAGGCACTGGAGAGAGAGGGCGCTCT 2150
Db 3936456 CCGCGACCGCGCTTCCTGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCGCGCGCG 3936397
OY 2151 GAGCGGACATCAGAGGCTCCAGAGGCTTCGCGCTCATCGACAGACCAAGTCCGCGGCA 2210
Db 3936396 ATTAGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCTGCGCGCGCG 3936337
OY 2211 GCGGACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTTTACGCTC 2270
Db 3936336 GCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936277
OY 2271 AGGGAACCGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2330
Db 3936276 TTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936217
OY 2331 GTACCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2390
Db 3936216 CCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936157
OY 2391 CTTCGACGCTGACTTCTCGAGAGACCAAGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 2438
Db 3936156 CTACCAACCGCGCTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936109

RESULT 4
AADI17184
ID AADI17184 standard; DNA: 65140 BP.
XX
AC AADI17184;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KM Polyketide synthase; PKS; macroide: nystatin; PKS gene cluster;
KM antifungal; antibiotic; nys1; ds.
XX
OS Streptomyces noursei.
XX
FH Key location/Qualifiers
FH CDS complement (1..1035)
FT /tag- a
FT /product- "NysD2 partial protein"
FT /note- "CDS does not include stop codon"
FT CDS complement (1035..2576)
FT /tag- b
FT /product- "NysD1 protein"
FT /note- "2806..6906"
FT CDS /tag- c
FT /product- "NysA protein"
FT /note- "6952..16530"
FT CDS /tag- d
FT /product- "NysB protein"
FT /note- "16550..49840"
FT CDS /tag- e
FT /product- "NysC protein"
FT /note- "50260..51015"
FT CDS /tag- f
FT /product- "NysE protein"
FT /note- "51405..54305"
FT CDS /tag- g
FT /product- "NysR1 protein"
```

```
FT CDS 54329..57190
FT /tag- h
FT /product- "NysR2 protein"
FT /note- "CDS does not include start codon"
FT CDS 57180..59963
FT /tag- i
FT /product- "NysR3 protein"
FT /note- "60415..61047"
FT CDS /tag- j
FT /product- "NysR4 (short) protein"
FT /note- "CDS does not include start codon"
FT CDS 61736..62497
FT /tag- k
FT /product- "NysR5 protein"
FT /note- "CDS does not include start codon"
FT CDS complement (62551..63615)
FT /tag- l
FT /product- "ORF2 protein"
FT /note- "CDS does not include start codon"
FT CDS 63765..64961
FT /tag- m
FT /product- "ORF1 protein"

W0200159126-A2.
XX
PN 16-AUG-2001.
XX
PD 08-FEB-2001; 2001MO-GB00509.
XX
PF 08-FEB-2000; 2000GB-0002840.
XX
PR 10-APR-2000; 2000GB-0008786.
XX
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STJFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEMSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (EJAE/) EJAEVIR E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX WPT: 2001-557614/62.
DR DR P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
PT
XX
PS Claim 2: Page 116-151; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macroide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
XX Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;
SQ
Query Match 3.2%; Score 87.2; DB 22; Length 65140;
Best Local Similarity 45.0%; Pred. No. 4.6e-09;
Matches 412; Conservative 0; Mismatches 498; Indels 6; Gaps 2;
```


Matches 412; Conservative 0; Mismatches 498; Indels 6; Gaps 2;

```

OY 1371 ACAGCATCGAGGATTCCTGTCAGGACGCGCTGTGACACCGGGGAGAGAGTAT 1430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80274 ACAGCGCCCGCGGGGACAGCGTGTCTGCTCGACTCCCTGGTCTCCGGAGATCCCG 80333
OY 1431 TGGCCGAGTCGAGAGACAGACGGCTTTCACCTTGGCCAGAGAAACAGCTGGGACAT 1490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80334 CCGACCGACCGCGGGCGCGGACGCTCCACCGGACGCTCCCTTTCACAGTGCAGTGA 80393
OY 1491 CCGCCGACAGAGGCTGTGGGCAACCGCCACATGGCCACCTTGACAGAGGCTGACTAC 1550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80394 CCGCCGCTCCACAGGGCGCCCGCGCCGACCGGCGACCGCTGCTGGCCCGCCGACC 80453
OY 1551 GGAACCTGGTCAAGAGCGTGTCCGCTCGGCAACCGCGCGGACCAAGAGTGGCCA 1610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80454 CGGACGCGCTTCGCGGACACCTTCGCGCCACCGGACATCCGAGCACCGCCCGCGAC 80513
OY 1611 TCTGCTCCAGCGCGCGCGGACGTCGAGATCGAGCGCGCGACGTCATGAGC 1670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80514 TGGCGCGCTTCGCGGACGCGGAGAGGCGCGCTCCCGGACGTCGTCACACCGCTCA 80573
OY 1671 CGCGCAAGGTGGCTTATTCGGCGGACGAGACGATCTCTGTGTCACCGCGCTGTCG 1730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80574 CCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 80633
OY 1731 GCGTGCAGCGCTTCGACAGTTCAGGAGGACGCTTGGCTTCGAGCGCTCCCGCGGCG 1790
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80634 TCGGCCAAGATGGCTGCGCGGACGACGCGTT---CGCGAGCGCGCGCTGTCTGTCA 80690
OY 1791 CCGTGCATGCGCTTCGAGCAAGAGAACACAGCTTCTTACGCGCGGCTCCGGATCGACT 1850
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80691 CCGCGCGCGCGCGCGCGAGGACCGGACCGCGCGCGCGCGCGCGCGCGCGCTGATCC 80750
OY 1851 TTACGTGAGCAAGGAGACGAGACGAGCTTTCACGCGCGGCGCGCGCGCGCGCGAG 1910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80751 CCGCGCGCGCGAGAGAACCGCGCGCTTTCGCTTCGACTCGCGCGCGCGCGCGCG 80810
OY 1911 GAGAGATTCGAGATATGCTGCTACCGGACACCGCGCGGCGAGCTTGTATGTCTGAC 1970
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80811 GCGCGCGCGCGAGAGACCTGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCG 80870
OY 1971 ACCTGCGCATATTCGCTTCACAGACTCGGCGGACGACTTGGCGCAAGTCTCCACCGCC 2030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80871 CCGTCCGCGGACGAGGTGACGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCGAAC 80930
OY 2031 TGACACACCTACAGATGCGCTGCGGTGGGTGAGGCTCGAAGTGAACCTGTATG 2090
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80931 CCACACCTGGAACCGCGGACGCGACGCTCTGATACCGCGCGCGCGCGCGCGCGCG 80990
OY 2091 CATTGCGGACCGCGCGCTTCAGGGGCTGCTTACGCGCACTGAGAGCGCGCGCTCT 2150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 80991 CCGTCCGCGCGCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81050
OY 2151 GAGCGGATTCAGGGGCTTCAGGGGCTTCGGC---TTCATCGACAGAGCAAGTTCGCG 2207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 81051 GCGCGCGCGCGCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81110
OY 2208 GCAAGCGGACGCGCGCGCGCGGCAAGTCTAGCTGGGACCAACGCGCGCGCGCTTTTAC 2267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 81111 CCAAGCTTCAGATGCGCGCGCTGCGAGCTCGCGGACCGCGCGCGCTGCGCGACTG 81170
OY 2268 CTCAGGAGACCGTGG 2283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 81171 GCACCGTCCCGCGCG 81186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6
AA064206 standard; cDNA; 1833 BP.

XX AA064206;
XX AC
XX

```

DT 18-NOV-1994 (first entry)
XX
DE snbr gene encoding enzyme in streptogramin biosynthetic pathway.
XX
KW Antibiotic; streptogramin; snaa; snab; snac; biosynthesis; enzyme;
KW biosynthetic pathway; Streptomyces pristinaespiralis; snba; snbr;
KW ds.
XX
OS Streptomyces pristinaespiralis.
XX
FH Key location/Qualifiers
FT CDS 103..1689
FT /tag= a
XX
PN FR2696189-A.
XX
PD 01-APR-1994.
XX
PE 25-SEP-1992; 92FR-0011441.
XX
PR 25-SEP-1992; 92FR-0011441.
XX
PA (RHON ) RHONE POULENC RORER SA.
XX
PI Blanc V, Blanche F, Crouzet J, Jacques N, Lacroix P;
PI Thibaut D, Zagorec M;
XX
DR WPI: 1994-128286/16.
DR P-PSDB; AAR54206.
XX
PT DNA involved in streptogramin antibiotic biosynthesis - for
PT prodn. or bio-conversion of streptogramin(s) or prodn. of
PT streptogramin intermediates, derivs. or hybrid antibiotics
XX
PS Claim 2; Page 62-64; 83pp; French.
XX
CC The snbr gene product is involved in the biosynthesis of
CC streptogramins, antibiotics active against Gram-positive bacteria.
CC The identification of the sequences encoding the enzymes involved
CC in the biosynthetic pathway means that they can be isolated and
CC manipulated. Mutant microorganisms in which a step in the
CC streptogramin biosynthetic pathway is blocked can be cultured to
CC produce streptogramin intermediates, which may later be converted
CC to streptogramin derivatives. Recombinant cells may also be used
CC for the bioconversion of streptogramins from one form to another or
CC for the production of hybrid antibiotics.
XX
SQ Sequence 1833 BP; 232 A; 802 C; 564 G; 235 T; 0 other;
XX
Query Match 2.98; Score 78.4; DB 15; Length 1833;
Best Local Similarity 46.18; Pred. No. 1.6e-07;
Matches 339; Conservative 0; Mismatches 391; Indels 6; Gaps 2;

```

```

OY 1523 GCCACCTCGAGAGCGCTGACTACGCGGAGACTCGGTCAAGAGCGTCCGCGTGGC 1582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 820 GCCATATCTTCGCGCGCGCGGACAGAGCTGAGACGCGCGCGCGCGCGCGCGCGCG 879
OY 1583 AACACCGCGCGGACGAAAGTGGCCATCTGTCGACGCGCGCGCGGAGAGTATC 1642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 880 GTGCGGCGCGCGCGCTCACCGGCTTCTCGCTGAGAACATGCAACCCCGCACCCATG 939
OY 1643 GACTACGCGCGCGGACGATGATGAGCGCGGACAGGTGACCTATTTGGCGGAGCGAG 1702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 940 CTCGACATGCGGCTTCTTACCGAGCGCGCTTCAACGCGCGCTGCGCGGAGTG---C 996
OY 1703 ACATCTCTGTCGAGCGCGCTGCGCGCGGAGCGCTGCAAGTTCAGGCGAGCTTT 1762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 997 TCGTCGTTGCGATGAGCGCGCGGCGCTTCTGCTGACCGACGACTCCAGTCTGCTC 1056
OY 1763 GCCTCGCTTCGAGCGCTGCGCGCGGCGCGCTGATCGCTCGGAGAAAGAGCAAGC 1822
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1057 GGCTAGACGCGCGCTGAGGCGCGGCGCTGCGACCGCGCGCACTGGCTTTGACGATGTC 1116
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
OY 1833 GTCCTTCAACGCCGCTCCGGATCGACCTTTTACGTCGACAGAGACCGGACAGCTTC 1882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1117 CTCACACTGGCGCGCTGGCGGCAAACTCTCGCGGCTGGCGACCGCCGCGAGCATC 1176
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1883 ACGCCCGGGCCCAACTGGGAGCGGAGGAGCATCCGGGATATCGCTGCTACCCGACC 1942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1177 GCCCTGGCGATGACACTGGCCCGCCCGCTTCACGCGCGGCGCTCGCGGATCGGGC 1236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1943 ACCGCGGCGACCTTTATGTCGACCGGACGATTCGCTTCGCTCCAGACTCGGGC 2002
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1237 CCGGAGCGCGGCTAGCGGGGATGCTCGCGGCTCTCTCTATGGCGCGGATCGGA 1296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2003 ACGACCTTTGGCCAAAGTCTCACCGCCCTGACCAACACCTACAGATGCGCTG---GGT 2059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1297 CTGGCGCATGCCCGCATGGCCACCGCCGATGCTCTCATCTCCCGCCCGCAAGCGCGG 1356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2060 GTGGGCTAGGCTCGAAGTGGAAACCTGTATGCTTCGACGCGGCGCCCTCAGAGGCTCCG 2119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1357 GCGGAGCGCGGCTCGACAGGACCCCTGACCGGATTCGCGGCGGAGTGGAGTGGGATC 1416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2120 CTCTACGCCAGTGGAGACAGCGGCGCTCTGAGGAGCATCCAGGCTCCAGGCGCTTC 2179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1417 CTGGGCGCGCTCTGCGGCTCCGCTTCGCTCCCAACTGCCCGCCATCAGCGGAC 1476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2180 GGCTCCATCGACAGCAAGTCCCGCGGAGCGGACGCGCGGCGCAAGTCTACGTG 2239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1477 GGTCTCTCTGACAGGACTGCTCGACGACACCCCAACAGCGGCGGAGTCTCACGAC 1536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2240 GGCACCAAGCGCGCGG 2255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1537 GCGTTGCGCGACGCGG 1552
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 7
AA199683
ID AA199683 standard: DNA: 4403765 BP.
XX AA199683:
AC AA199683:
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
FN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
DR WPI: 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ.
XX
PS Claim 4; SEQ ID NO 2; 3bp + Sequence Listing: English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen.
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
```

CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 2.6%; Score 71.4; DB 22; Length 4403765;

Best Local Similarity 45.2%; Pred. No. 4e-05;

Matches 508; Conservative 0; Mismatches 598; Indels 18; Gaps 4;

```
OY 1175 AACCACTTATGATGATGTGACGACGCAAGTACCGTCCGATGCTCATCAAGCGCTCGGC 1234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3941794 AGCAGCGGTGCGCGCGGACCAACGCGCTCCGCGCGCGCGGCGGCGGACCAAGGC 3941853
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1235 TGGATGATTGATCTCTGAGATTGACCCACGACGACACCTGCGCTTAAGGCAACC 1294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3941854 GCGCGCGGGGGTGGCTGGCATGACTTCAGCAACGCGACGAACGCGGCGGCGGCAAC 3941913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1295 GGAATGACAAATCTTTGGCGCGCAGCATCTCACCACTGGGAGACGCGCCCAATGTGCA 1354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3941914 GCGGCGTGGCGGCGACCGGGGGGAGCGGCGCAACGACGAGCACCGGCGCGGACCCC 3941973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1355 ATCCATCACTGGCAGAGCGGCAATCGAAGAAATCTCCGTTCAGAGACTGGCTCTGACCC 1414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3941974 GGCAGAAAGCGGCGACCGGCGGACCGGCGGCGGCGGCGGCGGCGGCGGTGAC 3942033
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1415 GCGGAGCGGAGCTATTGTCGCGCGGAGAGAGACGACCAACGCGTTCCTTCCAGCAGA 1474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3942034 GCGCGGCGCACTTACACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3942093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1475 AACGACCTTGGGAGCATCGCGGAGACGCGTCTGGGCAACGCCACATGGGCGCATCTGACG 1534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3942094 AAAG-----GCGGCTTAACACCGGAGCTCAGAGGCGCACGAGCGGCGGCGGCGC- 3942147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1535 AGCGTGCATACGCGCGGAGCTCGGTCAAGAGCGTCTCGCGGTGGGCAACCGCGGCG 1594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3942148 ACCGCGGCGACCGGCGGGAAGGCGCGCACGCGCGGCGGCGGCGGCGGCGGCGGCGG 3942207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1595 ACGCAACAGGTGGCATCTCTGTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3942208 ACCGCGGCGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3942267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1655 GACACGTCCATGACGCGGCGGACGCGTGGCTTATGGCGGCGGCGGCGGCGGCGGCGG 1714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3942268 GGCACCGGACGCGCAAGCGGCGGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGA 3942327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1715 TCGACCGCTCTGTCGCGGCGTGCAGGCGTGCAGTTCCAGGCGAGCTTCCCTCGTGG 1774
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3942328 GGAAGACAGCGGTCTCGGATTTGGGCGG-----CCAGCGCGGCTTGGCGGCGGCC 3942378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1775 AGCCTGCGCGCGGCGGCGGCTCATCGCTCGACAGAAAGCAACCAAGCTTCTTACGCGC 1834
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3942379 GCGCGCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3942438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1835 GCGCTCGGATGACCTTTTACGTCAGACGACCGGCGGAGCTTCAAGCGGCGGCGG 1894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3942439 GCGGCGGCGGCGGCGGCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3942498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1895 AAGCTGCGGAGCGGACGAGCATTCGGGATATCGCTGTACACCGACGCGGCGGCGG 1954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3942499 AAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3942558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1955 TTGTATGCTTCGACGCGAGCTGCGCATATTCGCTCCACAGACTCGGCGGCGGCGGCGG 2014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3942559 GGCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3942618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2015 CAAGTCTCCACCGCGCTGACCAACACTACAGATGCGCTGGTGGGCTCAGGCTCG 2074
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```

XX MO200071577-A1.
PN
XX
XX 30-NOV-2000.
PD
XX
XX 25-MAY-2000; 2000MO-US14462.
PE
XX
XX 25-MAY-1999; 99US-0318208.
PR 20-JUL-1999; 99US-0144882.
PR 10-AUG-1999; 99US-0147823.
PR 13-DEC-1999; 99US-0373658.
PR 22-DEC-1999; 99US-0171503.
PR 22-FEB-2000; 2000US-0183792.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (BETH ) SMITHKLINE BECHAM CORP.
PA (BETH ) BETH ISRAEL DEACONESS MEDICAL CENT.
PA (TRUE) IRUELA-ARISPE L.
PA (HAST) HASTINGS G A.
PA (RUBE) RUBEN S M.
PA (JONA) JONAK Z L.
PA (TRUL) TRULLI S H.
PA (FORN) FORMALD J A.
PA (TERR) TERRETT J A.
XX
XX Irueja-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
PI Formwald JA, Terrett JA;
XX
XX MPI; 2001-025136/03.
XX
XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
XX rheumatoid arthritis and psoriasis -
XX
XX Claim 7: Pages 661-663; 768pp: English.
XX
XX The present invention relates to human METH1 and METH2, (ME for
CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC The present sequence is an expressed sequence tag (EST) for METH. METH
CC can be used for inhibiting angiogenesis in an individual, and for
CC treating cancer, benign tumours, an ocular angiogenic disease,
CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC vasculogenesis, granulomas, hypertrophic scars, nonunion fractures,
CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC coronary collateral, cerebral collaterals, arteriovenous malformations,
CC ischaemic limb angiogenesis, Osler-Weber syndrome, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiodioma,
CC fibromuscular dysplasia, wound granulation, Crohn's disease or
CC atherosclerosis. METH can also be used in birth control. METH can also
CC be used in diagnostic methods for the prognosis of cancer.
XX
XX Sequence 3015 BP; 674 A; 1079 C; 831 G; 431 T; 0 other;
S0

```

```

Query Match 2.68; Score 69.8; DB 22; Length 3015;
Best Local Similarity 43.18; Pred. No. 1.4e-05;
Matches 396; Conservative 0; Mismatches 517; Indels 6; Gaps 1;

```

```

OY 1445 GACGACACGCGCTTACCTTGCAGCAAGAAAGCCTCGGCATCGCCGACAGCGTC 1504
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1478 GCCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1517
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1505 TGGCGAAGCGCCACATGCGCCACCTCGACGACGCTGACGACGCGGGAAGTGGTCAAG 1564
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1538 AGCGACGCTACACACGACGACGACGACGACGACGACGACGACGACGACGACG 1597
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1565 AGCGTCTCGCGCTGCGCAACCGCGCGGACGACGACGACGACGACGACGACGACG 1624
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1598 AGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1657
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1625 GCGCGACGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1684
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1658 GAGCGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1717
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

OY 1685 TAATCGCCGACGCGACACGACGACGACGACGACGACGACGACGACGACGACGACG 1744
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1718 AGCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1777
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1745 CAGTTCAGGGCAGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1804
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1778 ACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1837
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1805 GACAGAAGACCAACAGCTCTTCTACGCGCGCTCCGAGTTCGACCTTTTACGTACGACAG 1864
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1838 AGCGAGGCTACACACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1897
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1865 GACACGCGACGACGCTTTCACGCGCGGCGCCAAAGCTGGCGACCGCAGGAGATCCGGAT 1924
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1898 ACCACGACGACGACGCTTACACGACGACGACGACGACGACGACGACGACGACGACG 1957
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1925 ATGCGTCTGACCGGACGACGCGCGGCGGACGCTTGATGCTTCGACCGGCTGCGATATTC 1984
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1958 GACGCTACACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2017
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1985 CGCTCCAGACACTCGGCGACGACCTTTTGCCAAAGTCTCCACCGGCTGACACACCTAC 2044
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2018 ACCAGACGACGACGACGCTTACACGACGACGACGACGACGACGACGACGACGACGAC 2077
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 2045 CAGATCCCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2104
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2078 GCTACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 2105 CCGTCAAGGCGCTGCGCTTACGCGGAGTGA-----CAGCGGCGCTCTGAGAGGAC 2158
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2138 AGCAGCAGGCTTACCGGCTACCGACGACGACGACGACGACGACGACGACGACGACGACG 2197
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 2159 ATCCAGGCTCCAGAGGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 2218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2198 ACCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2257
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 2219 ACCGCGCGGCAAGCTTACGCTGCGGACCAAGCGCGCGGCTCTTTTACGCTGACGAGAC 2278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2258 AGCGACGCTACACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2317
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 2279 GTCGCGCGGCGACGCGCGGCGGACTTCCTGTCGACCAAGCAGACGACGACGACGACGACG 2338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2318 ACCAGCAGCAGCAGGAGGCTTACCGGCTACCGACGACGACGACGACGACGACGACGACG 2377
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 2339 TCGCGACGCTGACCA 2357
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2378 GAGGCTACCAACGACGACCA 2396
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 11

```

AI199682
ID AI199682 standard; DNA; 4411529 BP.
XX
XX AI199682;
AC
XX
XX 15-JAN-2002 (first entry)
DT
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
DE
XX
XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX US6294328-B1.
PN
XX
XX 25-SEP-2001.
PD
XX
XX 24-JUN-1998; 98US-0103840.
PF
XX
XX 24-JUN-1998; 98US-0103840.
PR
XX
XX (GENO-) INST GENOMIC RES.
PA

```

```
XX  Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
XX  WPI: 2001-647261/74.
DR
XX  Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX  determining the nucleotide sequence of the strain at positions in the
XX  genome corresponding to positions where M. tuberculosis strains CDC
XX  1551 and H37Rv differ.
XX
XX  Claim 3: SEQ ID NO 1; 3pp + Sequence Listing: English.
XX
XX  The invention relates to evaluating strain variation within and between
XX  different populations of the tuberculosis bacterial pathogen,
XX  Mycobacterium tuberculosis or related Mycobacterium by determining the
XX  nucleotide sequence of the first strain at positions in the complete
XX  sequence of the genome that correspond to positions that differ in the
XX  nucleotide sequences of M. tuberculosis strains CDC 1551 (A19683) and
XX  H37Rv (A19682). The method is useful for evaluating strain variation of
XX  M. tuberculosis and has valuable application in the fields of
XX  tuberculosis genetics, epidemiology, patient treatment and epidemic
XX  monitoring.
XX  Note: The sequence data for this patent did not form part of the printed
XX  specification, but was obtained in electronic format directly from USPTO
XX  at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
XX  Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other:
XX
XX  Query Match 2.6%; Score 69.4; DB 22; Length 4411529;
XX  Best Local Similarity 43.2%; Pred. No. 0.00011;
XX  Matches 503; Conservative 0; Mismatches 646; Indels 16; Gaps 3;
XX
XX  1367 GCAGAGCGCATGAGGAATTCCTCCAGAGACTGGCCTTCGACCCGCGCAAGACGAG 1426
XX  DB 335004 GCGCGCAGCGCCACGAGATCCCGCGGAGAGCCCTGCGCAACTCCGCTTCCGCGCC 335003
XX
XX  1427 CTAATTGCGCGAGTGGAGACGACAAAGGCTTCACCTTTCGACGAGAAAGCACTCGGG 1486
XX  DB 335064 GCGCGCGGCGGAGACCGTTCGCGCGCGCGCGCTTCCGCGCGCGCGCGCGCG 335123
XX
XX  1487 ACATCGCGCGCAGACGGTGTGGGCAACGCGCACATGGGCGACCTCGACGAGCGTGCATAC 1546
XX  DB 335124 GAAGCCGAATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335183
XX
XX  1547 GCGGGAATCTCGCTCAAGAGCTCGCTCCGCTCGCGCAACCGCGCGCGCGCGCGCGCG 1606
XX  DB 335184 GCGGCGCGCGCGCGCGCGCGCGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCG 335243
XX
XX  1607 GCATCTCTGCTCGAGCGCGCGCGCGCGAGCATGACATGACGCGCGCGCGCGCGCGCG 1666
XX  DB 335244 TCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335303
XX
XX  1667 AACGCGCGCGCAGTGGCTATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1726
XX  DB 335304 GTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335363
XX
XX  1727 TCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1786
XX  DB 335364 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335423
XX
XX  1787 GCGCGCGCGCTATGCTCGAGCAAGAAAGCAACAGCGCTTCTTACGCGCGCGCGCGCG 1846
XX  DB 335424 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335483
XX
XX  1847 ACCTTTACGTGACGAAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1906
XX  DB 335484 GAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335543
XX
XX  1907 GCAGGAGCATCCGCGGATATCGCTGCTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966
XX  DB 335544 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335603
XX
XX  1967 ACCGAGCTCGGCGATATTCGCGCTCACAGCATCGGCGCGCGCGCGCGCGCGCGCGCG 2026
```

```
DB 335604 TCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335663
XX
XX  2027 GCGCTGACCAACACTTACAGATCCGCTGCGGTGTGGGTCTAGGCTCGAATGGAACCTG 2086
XX  DB 335664 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335720
XX
XX  2087 TATGCGCTTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2146
XX  DB 335721 CATGGCGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335780
XX
XX  2147 TCCGTGACGAGCATCCAGGCTCCAGGCGCTTCCGCTTCATGACAGCAACCAAGTCCGC 2206
XX  DB 335781 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335840
XX
XX  2207 GCGACGCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2266
XX  DB 335841 GTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335900
XX
XX  2267 GCTCAGGAGAACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2316
XX  DB 335901 GATTAAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335960
XX
XX  2317 GCAGAGCAGCAGCAGTACTCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2376
XX  DB 335961 GCGGATCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336020
XX
XX  2377 ATCCAGACCGCGCGCGCTGACGCGTGAATCTGCGAGAGACCACTGGCGCGCGCGCG 2433
XX  DB 336021 CGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336080
XX
XX  2434 CAGCGGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2493
XX  DB 336081 GAACGTGGCGATGAAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336140
XX
XX  2494 GACGCAAGTGTGTGGCGCGCGTATGTC 2518
XX  DB 336141 GCGGAAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336165
XX
XX  RESULT 12
XX  AAH74540
XX  ID AAH74540 standard: cDNA: 1107 BP.
XX
XX  AAH74540;
XX  AC
XX  AC 15-OCT-2001 (first entry)
XX  DT
XX  DE Nucleotide sequence of an endochitinase polypeptide.
XX  KN Chitinobiosidase; chitinolytic enzyme; early flowering; ss.
XX  OS Streptomyces albidoflavus.
XX  PN WO200146387-A1.
XX  PD 28-JUN-2001.
XX  PF 22-DEC-2000; 2000WO-US35238.
XX  PR 23-DEC-1999; 99US-0172003.
XX  PA (CORR ) CORNELL RES FOUND INC.
XX  PI Broadway RM, Gongora CE;
XX  DR WPI: 2001-496661/54.
XX
XX  Promoting early flowering, yield in plants and reducing plant size,
XX  involves growing transgenic plant or plant seed transformed with DNA
XX  molecule encoding chitinolytic enzyme, or applying the enzyme to the
XX  plant
```


PS Disclosure: Page 18-19; 87pp; English.

XX The present sequence encodes an endochitinase polypeptide. The

CC polypeptide is a chitinolytic enzyme. The chitinolytic polynucleotide

CC is used to produce transgenic early flowering plants. The chitinolytic

CC polynucleotides and polypeptides are useful for promoting early

CC flowering in plants, promoting yield from plants and reducing plant

CC size.

XX

Sequence 1107 BP; 202 A; 427 C; 331 G; 147 T; 0 other;

SO

Query Match 2.5%; Score 66.8; DB 22; Length 1107;

Best Local Similarity 44.5%; Pred. No. 4.9e-05;

Matches 405; Conservative 0; Mismatches 492; Indels 13; Gaps 3;

1515 CCACATGGGCGCCACCTCGACGACGCTGACCTGCGGGAACCTGCGTAAGAGCGTCGCC 1574

131 CCATCGGTGACAGCTTGGCCCGCTACGACAGCGGTACACCGCGGAGTGGTCGACG 190

1575 GCGTCGGCAACCCCGCGCAGCAGAGTGGCCATCTCTCCGACGGCGCGCGACGT 1634

191 GCGTCGGCAACCTGGGACAGCGCGGTGGCGGCACTTCAACAGCTCCGCAAGCTCA 250

1635 GCAGCATGCACTACGCGCGCAGACGTCATGAACGGCGGACGCGTATTCGGCGC 1694

251 AGGCGAAGTACCCGACATCAAGGTCCTTGGTCTTGGCGGCTGACCTGTCCGGCG 310

1695 ACGGCGACAGATCTCTGGTCGACCGCTGCGGCGTGCACGCGGCGAGTTCCAGG 1754

311 GCTTACACGACGCGCTGACAGAACCGCGCTTTCGCGCACTCTGCGACAGACTGGTCG 370

1755 GCACCTTGGCTCCGCTCGAGCTGCGCGCGCGCGCTGTCATCGCTCGGCAAGA 1814

371 AGACCGCGCGTGGCGGAGCTTTCGAGCGGATCGACCTGAGTGGAGTACCGGAAG 430

1815 CCACAGCGCTTCTACGCGCGCTCCGATGACCTTTTACGTAGCAAGACACCGGCA 1874

431 CTTGCGGCTGAGCTGAGACGCTCGGTCGCGCGCTGAGAACATGTTCCAGGCGA 490

1875 GCAGCTTACGCGCGCGCGCGCAAGCTGGGACGCGAGGAGGATCCGGATATCGCTCTC 1934

491 TCGCGCGCGAGTTGGGACGAGCTGGTCAACCGCGCGCATACGCGCGCGCGAGCTCCG 550

1935 ACCGACACCGCGCGGACGTTGATCTCTGACGACGTCGCGATATTCGCTCCACAG 1994

551 CGCGCAAGCTCTGACGCGCGCGGACTACG---CGGCGCGCGCGCACTACTGACTGATCA 607

1995 ACTGGGCGACGACCTTGGGCAAGTCTCCACCGCGCTGACCAACACTACAGATTCGCC 2054

608 ACGTATGATGACCTGACTTCTTCCGCGCTGGGACAGACCGCGCGCGCGCGCACT 667

2055 TGGGTGTGGGCTCAGGCTCGAAGTGAACCTGTATGCTTGGCGACCGCGCGCTCAGGG 2114

668 CGGCGCTGAACTCTTACAGGGGATCCCAAGGCGGACTTCACTGGCGCGCGCGCAATCG 727

2115 CTGGCTCTTACGCAAGTGAAGACAGCGCGCTCTCTGAGCGGACATCAGGCTCCAGG 2174

728 CCAAGCTCAAGGGGAAAGGGGCTCCCGCGAGCAAGCTCTGCTGCGATGCGGCTTTCAG 787

2175 GCTTGGGCTCATGAGACAGCAACAGTGGCGGCGGCGGAGACGCGCGGCGGAG--- 2231

788 GCGCGCTGAGACCGGCTCAACCAAGAGCGCGCGGCGGCGCGCGCGCGCGCGCA 847

2232 -----TCTAGTGGGCGACCAACGCGCGGCGCTTTTATGCGTCAAGGAACGTCGCG 2285

848 CCGGACACTTACAGAGGCGGCGATCGAGATCAAGATCTCAAGAACACTGCGCGCGCA 907

2286 GCGGACAGGCGGCGGACTTCTCTGTCAGCAAGACAGACAGACAGTACTTTCGCGCA 2345

908 CCGGACCGCTGGGCGGCGGCTGACCAAGTCCGCGACCACTGCTGGAGCTACGACA 967

2346 GCTGACACACAGCTGAGTCA--GCGTGTATTCAGAGACCGGCGGCTTGGAGCGTCACT 2404

DB 968 CCCCCGCCACCATCAAGACCAAGATGACTGGGCCAAGAGACAGGCGCTCGCGCGCT 1027

QY 2405 TCGTCGAGA 2414

DB 1028 TCTTCTGGA 1037

RESULT 13

AA206825

ID AA206825 standard; DNA: 2712 BP.

XX

AC AA206825;

XX

DT 09-NOV-1999 (first entry)

XX

DE Streptomyces albidoflavus endochitinase DNA.

XX

KW Chitin: chitinolytic enzyme; insect; fungus; resistance;

KM transgenic plant; insecticide; fungicide; ds.

XX

OS Streptomyces albidoflavus.

XX

FH Key location/Qualifiers

FT mat_peptide 1355..2662

FT /*lag- a

FT /product= "Mature endochitinase"

FT /function= "Randomly cleaves chitin molecule internally"

XX

PN M09942594-A1.

XX

PD 26-AUG-1999.

XX

PF 12-FEB-1999; 99MO-US03176.

XX

PR 18-FEB-1998; 98US-0025691.

XX

PA (CORR) CORNELL RES FOUND INC.

XX

PI Broadway RM, Harman GE;

XX

DR WPI: 1999-527480/44.

XX

DR P-PSDB; AAY33722.

XX

PT New nucleic acid encoding chitinolytic enzyme active under alkaline

PT conditions, used to generate plants resistant to insects and fungi

PT and to produce recombinant enzyme

XX

PS Claim 11: Page 13-15; 90pp; English.

XX

CC This sequence represents Streptomyces albidoflavus endochitinase

CC DNA. Endochitinase hydrolyses chitin, an insoluble linear

CC beta-1,4-linked polymer of N-acetyl-beta-D-glucosamine found in all

CC arthropods, yeast, most fungi, and some stages of nematodes. There are

CC three types of chitinolytic enzyme activity: glucosaminidase, which

CC cleaves monomeric units from the terminal end of chitin; chitinobiosidase,

CC which cleaves dimeric units from chitin terminal ends; and

CC endochitinase, which randomly cleaves the chitin molecule internally.

CC Endochitinase and chitinobiosidase (AAY33721) from S. albidoflavus are

CC active under alkaline conditions. This makes them particularly effective

CC in controlling insects, as these enzymes can be ingested by insects and

CC can then attack them by degrading their chitin-containing, alkaline

CC digestive tracts. These chitinolytic enzymes can be applied directly to a

CC plant to act as an insecticide or fungicide, or non-pathogenic bacteria

CC transformed with the DNA encoding either or both enzymes can be applied

CC to a plant. Alternatively, transgenic plants can be produced which

CC express either or both enzymes in some or all of their tissues. A wide

CC variety of commercially important crops could benefit from use of these

CC enzymes e.g. sugar cane, potato, cotton and coffee.

XX

SO Sequence 2712 BP; 459 A; 1094 C; 830 G; 329 T; 0 other;

Query Match 2.5%; Score 66.8; DB 20; Length 2712;

Best Local Similarity 44.5%; Pred. No. 6.1e-05;

Matches	405: Conservative	0: Mismatches	492: Indels	13: Gaps	3:
OY	1515 CCACATGGGCAACCTCTGACAGAGGCTGACTACCGCGGGAACCTGCTCAAGAGCGTGTCC	1574			
DB	1698 CCATCGGTGACAGCTTCCGCGCTTACGACAGGCTACCCGCGGATGCGTGTGACG	1757			
OY	1575 GCGTCGGCAACCCGCGGACGCAACAGTGGCCATCTCTGCTCGAGCGGCGGCGACGT	1634			
DB	1758 GCGTCGCCACAGCTGGAGCAGACCGCTCGCGGCAACTTCAACGAGCTCCGACACTCA	1817			
OY	1635 GAGCATTCACATGAGCGGCGGACAGCTCATGACAGCGGCGGACGTGGCTATTGCGCG	1694			
DB	1818 AGGCCAAGTACCGGACATCAAGTCTCTGTCTTCCGCGGCTGAGCTGTGTCGGCG	1877			
OY	1695 ACGGCGACAGATCTCTGTGACCGCTCTGCGGCGTGCAGCGCTGCGCATTTCCAG	1754			
DB	1878 GCTTACCGACCGCGTGAAGAACCGCGCTTCCGCAAGTCTCTGACAGACTGTG	1937			
OY	1755 GCAGCTTTCCTCCGCTCTGAGCTTCCGCGGCGGCGCTCATCGCTCGGACAGAGA	1814			
DB	1938 AGGACCGCGCTGGCGGACGCTTCTGAGCGGATGACCTCGAGTGGAGTACCGAGC	1997			
OY	1815 CCAACAGCCTTCTCTAGCGCGGCTCCGATCGACCTTTTACGTACAGAGACCGGCA	1874			
DB	1998 CCGCGGCTTCAAGCTGAGACGCTTCCGCTCCGCGCTGAAGAACATGTCTCAGGCA	2057			
OY	1875 GCAGCTTTCAGCGCGGCGGCAAGCTGGGACGCGAGGACGATCCGGATATGCTGCTC	1934			
DB	2058 TCGCGCCCGAGTTTCCGACCGACGCTGTGACCGCGCATCACCGCGGACGCTCCG	2117			
OY	1935 ACCGACCAACCGGCGGACGTTGTATGTCTGACGCGAGCTGGCATATTCGCTCCAG	1994			
DB	2118 GCGGCAAGCTCGACGCGCGGCTGACTAGC---CGGCGCGCGCGCTGACTTGCATGTG	2174			
OY	1995 ACCTGGGCGAGACCTTGGCGAAGTCTCCACCGCGCTGACCAACACTACCAATGGCC	2054			
DB	2175 ACCTGTATGAGTACGACTTCTTGGCGCTGGGACAAAGCCGCGACCGCGCCCACT	2234			
OY	2055 TGGGTGTGGGCTCTAGGCTCGAAGCTGAACTGTATGCTTTCGCGACCGCGCTCAGGG	2114			
DB	2235 CGGCTTGAATCTCTACAGCGGCTCCCAAGCGGCTTCCATCTGGCGCGCGCATG	2294			
OY	2115 CTGCGCTTACGCACTGAGAGACGCGGCTCTGAGAGGACATCAGGCTCCAGG	2174			
DB	2295 CCAAGCTCAAGGGGAGGCGGTCCCGCGAGCAAGCTCTGCTGGCATGCGCTTCTAG	2354			
OY	2175 GCTTGGGCTCATCGACAGACCAAGTGGCGCGGACGCGGCGGCGAG---2231				
DB	2355 GCGCGGCTGGACCGGCTACCCAGAGCGCCCGGCGGCGCACCGCGCGCGCA	2414			
OY	2232 -----TCTACGTGGGCAACAGCGCGGCGCTCTTTTACGCTCAGGGAACGTCGGG	2285			
DB	2415 CCGGCACTACGAGGGGCGGATGAGAGCTACAAAGTCTCTCAAGAACACTGCGCGCA	2474			
OY	2286 GCGGCAACGGGAGGACTTCTGTGCAACCAAGCAGACAGCAGTACCTTTCGCGCA	2345			
DB	2475 CCGGCAACGGGAGGACTTCTGTGCAACCAAGCAGACAGCAGTACCTTTCGCGCA	2534			
OY	2346 GCTGAGACCAAGCTGAGAGTGA--GCGTTGTATCCACAGCCGCGGCTTTCAGAGGAGC	2404			
DB	2535 CCGGCAACGGGAGGACTTCTGTGCAACCAAGCAGACAGTACCTTTCGCGCA	2594			
OY	2405 TCGTCGAGA 2414				
DB	2595 TCTTGTGGGA 2604				

RESULT 14
AAH74538
ID AAH74538 standard; DNA; 2712 BP.
XX
AC AAH74538;
XX

DT	15-OCT-2001 (first entry)	
XX		
DE	Nucleotide sequence of an endochitinase polypeptide.	
XX		
KW	Endochitinase; chitinolytic enzyme; early flowering; ss.	
XX		
OS	Streptomyces albidoflavus.	
XX		
FT	Key	
FT	Location/Qualifiers	
FT	1535..2665	
FT	/*tag= a	
FT	/product= "Endochitinase"	
PN	MO200146387-A1.	
XX		
PD	28-JUN-2001.	
XX		
PF	22-DEC-2000; 2000WO-US35238.	
XX		
PR	23-DEC-1999; 99US-0172003.	
XX		
PA	(CORR) CORNELL RES FOUND INC.	
XX		
PI	Broadway RM, Gongora CE;	
XX		
DR	* WPI: 2001-496661/54.	
DR	P-PSDB; AAG63553.	
XX		
PT	Promoting early flowering, yield in plants and reducing plant size,	
PT	involves growing transgenic plant or plant seed transformed with DNA	
PT	molecule encoding chitinolytic enzyme, or applying the enzyme to the	
XX	plant	
PS	Claim 11; Page 15-16; 87pp; English.	
XX		
CC	The present sequence encodes an endochitinase polypeptide. The	
CC	polypeptide has a molecular mass of 45 kD and an isoelectric point	
CC	of less than 6.5. The polypeptide is a chitinolytic enzyme. The	
CC	chitinolytic polynucleotide is used to produce transgenic early	
CC	flowering plants. The chitinolytic polynucleotides and polypeptides	
CC	are useful for promoting early flowering in plants, promoting yield	
CC	from plants and reducing plant size.	
SQ	Sequence 2712 BP; 459 A; 1094 C; 830 G; 329 T; 0 other;	
XX		
Query Match	2.5%; Score 66.8; DB 22; Length 2712;	
Best Local Similarity	44.5%; Pred. NO. 6.1e-05;	
Matches 405; Conservative	0; Mismatches 492; Indels 13; Gaps 3;	
OY	1515 CCACATGGGCAACCTCTGACAGAGGCTGACTACCGCGGGAACCTGCTCAAGAGCGTGTCC	1574
DB	1698 CCATCGGTGACAGCTTCCGCGCTTACGACAGGCTACACCGCGCGGATGCGTGTGACG	1757
OY	1575 GCGTCGGCAACCCGCGGACGCAACAGTGGCCATCTCTGCTCGAGCGGCGGCGACGT	1634
DB	1758 GCGTCGCCACAGCTGGAGCAGACCGCTCGCGGCAACTTCAACGAGCTCCGACACTCA	1817
OY	1635 GAGCATTCACATGAGCGGCGGACAGCTCATGACAGCGGCGGACGTGGCTATTGCGCG	1694
DB	1818 AGGCCAAGTACCGGACATCAAGTCTCTGTCTTCCGCGGCTGAGACTGTGTCGGCG	1877
OY	1695 ACGGCGACAGATCTCTGTGACCGCTCTGCGGCGTGCAGCGCTGCGCATTTCCAG	1754
DB	1878 GCTTACCGACCGCGTGAAGAACCGCGCTTCCGCAAGTCTCTGACAGACTGTG	1937
OY	1755 GCAGCTTTCCTCCGCTCTGAGCTTCCGCGGCGGCGCTCATCGCTCGGACAGAGA	1814
DB	1938 AGGACCGCGCTGGCGGACGCTTCTGAGCGGATGACCTCGAGTGGAGTACCGGACG	1997
OY	1815 CCAACAGGCTCTCTAGCGCGGCTCCGATCGACCTTTTACGTACAGAGACCGGCA	1874
DB	1998 CCGTGGCGCTTCAAGCTGAGACGCTTCCGCTCCGCGCTGAAGAACATGTCTCAGGCA	2057

Db 789 CAGCAGCAGCGTCGGCTACCCGGGCAAGTACCCGAGCGTCATCGCCGTCGGCCGCTGA 848
QY 2074 GAACGTGAACCTGTATGCTTCGGCAACCGCCCGTCAGGGGCTCGGCTCTACGCCAGTGG 2133
Db 849 CAGCAGCAACCGACGGCGCGAGCTTCAAGCAGC--GTCGGCCCGGAGCTGGACGTCAATGGC 905
QY 2134 AGACAGCGCGGCTCTCTGAGGACATCCAGGGCTCCAGGGCTTCGGCTCCATCGACAG 2193
Db 906 CCGGGCGTCAGCATCCAGAGCACCTCCCGGGCAGCAACAGTACGGCCCTACAGCGGCAC 965
QY 2194 CACCAAGTCCGCGGAGCGGCAAGCACCAGGCAAGTCTACGTGGGCACCAAC 2248
Db 966 CAGCATGGCCAGCCCGCACGTCGCGGGCGCGCTGCACATCCTCAGCAAGCAC 1020

Search completed: May 4, 2003, 10:58:00
Job time : 14533.3 secs

THIS PAGE BLANK (USP10)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 2003, 23:52:57 ; Search time 62.7339 Seconds
(without alignments)
13247.929 Million cell updates/sec

Title: US-10-026-994-1
Perfect score: 2710
Sequence: 1 ccacgcgtccgcagcagtc.....ctattgaaaaaaaaaaaaa 2710

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCRTS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfilltest1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	3.2	4403765	4 US-09-103-840A-2	Sequence 2, Appli
2	76.8	2.8	1833	2 US-08-403-852D-6	Sequence 6, Appli
3	76.8	2.8	1833	3 US-08-510-646B-6	Sequence 6, Appli
4	76.8	2.8	1833	4 US-09-231-818-6	Sequence 6, Appli
5	71.4	2.6	4403765	4 US-09-103-840A-2	Sequence 2, Appli
6	70	2.6	2793	1 US-08-209-747-1	Sequence 1, Appli
7	70	2.6	2793	1 US-08-458-298-1	Sequence 1, Appli
8	69.4	2.6	4411529	4 US-09-103-840A-1	Sequence 1, Appli
9	68.8	2.5	2712	3 US-09-025-691-4	Sequence 4, Appli
10	65.4	2.4	1140	3 US-09-023-173-4	Sequence 4, Appli
11	60.2	2.2	2810	4 US-09-408-647A-1	Sequence 1, Appli
12	60	2.2	4411529	4 US-09-103-840A-1	Sequence 1, Appli
13	59.6	2.2	2367	4 US-09-056-556-201	Sequence 201, App
14	58.6	2.2	2367	4 US-09-072-596-196	Sequence 196, App
15	58.4	2.2	2249	3 US-08-814-052-19	Sequence 19, Appl
16	58.2	2.1	1656	4 US-09-385-028-14	Sequence 14, Appl
17	58.2	2.1	15079	4 US-09-385-028-14	Sequence 14, Appl
18	57.4	2.1	2064	4 US-08-343-428-1	Sequence 1, Appli
19	56	2.1	1248	4 US-09-105-537-7	Sequence 7, Appli
20	56	2.1	5970	3 US-09-320-878-21	Sequence 21, Appli
21	56	2.1	13613	4 US-09-105-537-3	Sequence 3, Appli
22	55.8	2.1	5392	4 US-08-403-852D-1	Sequence 1, Appli
23	55.8	2.1	5392	4 US-08-510-646B-1	Sequence 1, Appli
24	55.8	2.1	5392	4 US-09-231-818-1	Sequence 1, Appli
25	55.8	2.1	68750	3 US-09-335-409-1	Sequence 1, Appli
26	55.8	2.1	68750	4 US-09-568-102-1	Sequence 1, Appli
27	55.8	2.1	68750	4 US-09-567-969-1	Sequence 1, Appli

C 28	55.8	2.1	68750	4 US-09-568-480-1	Sequence 1, Appli
C 29	55.8	2.1	68750	4 US-09-568-486-1	Sequence 1, Appli
C 30	55.8	2.1	68750	4 US-09-568-472-1	Sequence 1, Appli
C 31	55.8	2.1	68750	4 US-09-567-899-1	Sequence 1, Appli
C 32	55.8	2.1	71989	4 US-09-443-501A-2	Sequence 2, Appli
C 33	55.4	2.0	2214	3 US-08-864-038A-1	Sequence 1, Appli
C 34	55.4	2.0	3331	3 US-08-864-038A-2	Sequence 1, Appli
C 35	55.4	2.0	3331	3 US-08-864-038A-4	Sequence 4, Appli
36	55	2.0	1896	4 US-09-311-626B-15	Sequence 15, Appli
37	55	2.0	3300	1 US-08-194-290-6	Sequence 6, Appli
38	55	2.0	3300	2 US-08-614-377A-6	Sequence 6, Appli
39	55	2.0	3300	4 US-09-142-648B-6	Sequence 6, Appli
40	54.8	2.0	28958	1 US-08-258-261B-6	Sequence 6, Appli
41	54.8	2.0	28958	1 US-08-456-837-6	Sequence 6, Appli
42	54.8	2.0	28958	1 US-08-457-342-6	Sequence 6, Appli
43	54.8	2.0	28958	1 US-08-457-666A-6	Sequence 6, Appli
44	54.8	2.0	28958	1 US-08-458-076A-6	Sequence 6, Appli
45	54.8	2.0	28958	1 US-08-764-233A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-103-840A-2/C
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match
Best Local Similarity 43.0%; Score 88; DB 4; Length 4403765;
Matches 433; Conservative 0; Mismatches 575; Indels 0; Gaps 0;

QY 1431 TGGCGCGAGTTCGAGAGCAAGAGGCTTCACCTTTCAGAGAGAAAGACCTCGGACAT 1490
DB 3937116 TGGCGCGAGTTCGAGAGCAAGAGGCTTCACCTTTCAGAGAGAAAGACCTCGGACAT 3937057
QY 1491 CGCGCGAGAGCGGTGGAGAGCCACATGGGCGACCTCGACGAGCGTGCATACGCGG 1550
DB 3937056 CGAGCGTTCACCGCTTGTCCACCGAGCGGGGCGGACCGCGCGGCGTGGG 3936997
QY 1551 GGAAGTGGTCAAGAGCGTGTCCGCGTCCGAGCAACCGCGGCGAGCAAGAGTGGCCA 1610
DB 3936996 TGGCGCGAGTTCGAGAGCAAGAGGCTTCACCTTTCAGAGAGAAAGACCTCGGACAT 3936937
QY 1611 TGTGCTCCAGCGCGCGCGAGCGTGCATGACATACGAGCGGCGACACCTTCATACAG 1670
DB 3936936 CCACCGGTGGCGCTTGTTCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936877
QY 1671 GGGCGAGGTGGCGCTTGTTCGCGCGAGCGGCGAGCGATGCTGTGGAGCGGCGTGGCG 1730
DB 3936876 GCCCGCGGTGGCGCGCGCGCGGTGATATTCGCGCGCGCGCGCGCGCGCGCGCGCG 3936817

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/403,852D
  FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: PCT/FR 93/00923
  FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: FR 92/11441
  FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
  NAME: Meyers, Kenneth J
  REGISTRATION NUMBER: 25,146
  REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (202) 408-4000
  TELEFAX: (202) 408-4400
  INFORMATION FOR SEQ ID NO: 6:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 1833 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
      MOLECULE TYPE: cDNA
      HYPOTHETICAL: NO
      ANTI-SENSE: NO
      ORIGINAL SOURCE:
        ORGANISM: S.pristinaespiralis
      FEATURE:
        NAME/KEY: CDS
        LOCATION: 103..1689
US-08-403-852D-6

Query Match      2.8%  Score 76.8;  DB 2;  Length 1833;
Best Local Similarity 45.9%;  Pred. NO. 7.4e-09;
Matches 338;  Conservative 0;  Mismatches 392;  Indels 6;  Gaps 2.

QY 1523 GCCACCTCGAGACGGGTGGACACTGACCGCCGGAGACACTCGGTCAAGAGGCTGTCGCCGTCGCCG 1582
Db 820 GCCATCATCTCCGCGCCCGGACACCGCTTGAGAGCGCCCGCCGACGCTCTCTCGCGCGCCG 879
QY 1583 AACACCGCCGCGACGCAACAGGTGGCCATCTTGCTCCGACGGCGCGCGACGAGTCGACATC 1642
Db 880 GTCCGGCGCCGCGCGCTCACCGCGCTTGCGTCGCTGGAGAACCTGCACACCCCGACCCCGCATG 939
QY 1643 GACTACGGCGCCGACACTGCATGAAACGGCGGCGACAGGTGGCTATTTCGGCCGACCGCGAC 1702
Db 940 CTCGCATGAGGCGTTCTTCACCGACCGCGCGCTTCAACGGCGCGCTCCGCGCGAGTG--C 996
QY 1703 AGCATCTCTGGGTGCACGCGCTGTCGCGAGCGTCAGCGCTCCAGTTCACAGGGACGTTT 1762
Db 997 TCGTGTTCGGGATGAGCGCGCTGCTCTTCTTCGCTACCCAGACACTCCAATCGTCTTC 1056
QY 1763 GCGTCCGTCGTGAGCTCTCCGCGCGCGCGCGCTCATGCGCTTCGGACAGAAGACCAACAGC 1822
Db 1057 GCGTTCGACGCGCTTCACAGCGCGCGCTTCGCGACCGCGCACTGGCTTTCGATCGTCGCC 1116
QY 1823 GTCCTTCTACGGCGCGCTCGGAGTGCACCTTTTACGTACGACAGGACACCGCGGACGAGCTTC 1882
Db 1117 CTCACACTGGCGCGGCTTCGGCGGAAACTCTTCGCGCGCGCTGGGACCGCGCCGACGATC 1176
QY 1883 ACCGCGCGCCCAACTGCGCGACGCCAGCGAGATCCGGGATATCGCTTCACCGGACC 1942
Db 1177 GCCCGTGGCATGACACTGCTGGCGCGCGCGCCCTACAGCGCGGTGGCGGTGGCGGATCGGCG 1236
QY 1943 ACCGCGGCGACGTTTATGTCTGCACCGACGTCGCGCATATTCGCGCTCCACAGACTCGGCG 2002
Db 1237 CCGGACCGCGCGCTACGGGCGCATGCTGCTCGCGCGCGCTGCTCTCTATGAGGCGCGGCGATCGCA 1296

```



```

US-09-231-818-6
: Sequence 6, Application US/09231818
: Patent No. 6171846
: GENERAL INFORMATION:
: APPLICANT: Blanc, Veronique
: APPLICANT: Bianche, Francis
: APPLICANT: Crouzet, Joel
: APPLICANT: Jacques, Nathalie
: APPLICANT: Lacroix, Patricia
: APPLICANT: Thibaut, Denis
: APPLICANT: Zagorec, Monique
: APPLICANT: Debussche, Laurent
: TITLE OF INVENTION: Polypeptides Involved In The
: TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: Flanagan, Henderson, Farabow, Garrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/231,818
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/403,852
: FILING DATE: 10-MAY-1995
: APPLICATION NUMBER: PCT/FR 93/00923
: FILING DATE: 25-SEP-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 92/11441
: FILING DATE: 25-SEP-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Meyers, Kenneth J.
: REGISTRATION NUMBER: 25,146
: REFERENCE/DOCKET NUMBER: 03806, 0054-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1833 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: S.pristinaespiralis
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 103..1689
: US-09-231-818-6

Query Match 2.8% Score 76.8; DB 4; Length 1833;
Best Local Similarity 45.9%; Pred. No. 7,4e-09;
Matches 338; Conservative 0; Mismatches 392; Indels 6; Gaps 2

Oy 1523 GCACCTCGACGACGGTGCACACTACGCGGGAAGTCCGTCGAAGAGCTCGTCGCGCGGCG 1582
Db 820 GCATCATCTCCGGCGCCGACACAGCGCTGAGAGCGCCCGCAGAGTCTCTTCTGCGGCGCC 879
Oy 1583 AACACCGCCGCGACCAACAGGTGGCCATCTCTGTCGACGAGCGGCGCGACGTGAGACATC 1642

```

Db 880 GTGGGGGGGGGGGGGCTACCGGCTTGCTGCGCTGGGAACTGCACACCCCAACCCATTG 939

Qy 1643 GACTACGGGCCCGAAGACAGTCCATTAAAGGGGGCAGCGTGGCTTAATTCGGCCGAC 1702

Db 940 CTCGCACATGAGGGCTTCTTACCCGACCGGGGGCTTAAACGGGGCGTGGCGGGAGTG---C 996

Qy 1703 ACGAATCCTCTGTGACACCGGCTCGTCCGGGCGGTGCAGGGCTCGAGTTCAGAGGACAGTTT 1762

Db 997 TCGTGTTTGGGATAGCGCGGCTCGTCCCTCTTCTGCTCAACCCAGACCTCCAACTGTCCTC 1056

Qy 1763 GCGTCCGTCTGAGACTGACCCGGGGGGCGGTCAATCGCTTCGGACAAGAACAACAGC 1822

Db 1057 GGCCTACGACCGCTTCGACAGCGCGGCTGGGACCGCGCCACTGGCTTTGAGAGATGTCGGC 1116

Qy 1823 GTCCTTACGGCGGGCTCGGATGACGTTTACGTCAAGCAAGAACACCGGACAGCTTC 1882

Db 1117 CTCAACTGGGCGGGCTTCGGGCGCAAACTCCTGCGCGGCTGGGACCGCGCGAGATC 1176

Qy 1883 ACGCGCGGGCCCAAGCTGGGCGAGCGCAAGGACGATCCGGGATATCGCTCAACCGGACC 1942

Db 1177 GCCCGGCGATGACACTCTGCGCGCGGCTCAAGCGGGGGCGTGGCGGATGGGG 1236

Qy 1943 ACCGGGGGACAGTGTATGTCTCAGCCGAGTGGGATATTCCGCTCAAGACTGGGC 2002

Db 1237 CCGGACGCGGGCTACGGGGGCGATCTGCGCGGCTGCTCTCATAGGCGGGGATACGA 1296

Qy 2003 ACGACCTTTGGCCCAAGTGTCCACGCGGCTGCAACCAACTACAGATGCGCTG---GGT 2059

Db 1297 CTGGGCAATGCCCGGCACTGGCCACCGCGGTATGTCTTCATCCCGCGCGCAAGCGCGG 1356

Qy 2060 GTGGGCTAGGCTCTGAACTGGAACCTGTATGCTTTCGGCAACCGGCGCGTCCGCTCGC 2119

Db 1357 GCGCGAGCGGGCGTGGAGGGGACCCCTACGGAAGTTTCGGGGGAGACTGGAGTGGGATC 1416

Qy 2120 CTCTACGCGCACTGGAGAGACGGGGCGCTCCGAGACGGAACATCCAGGGGCTCCAGGGCTTC 2179

Db 1417 CTCGGCGCGTCTTCGCTCCGCTTTCGCTTCCCACTGCCCGCGCATACCGGACCTC 1476

Qy 2180 GAGTCATCGACAGACACCAAGTGGCGGCGAGCGGACGACCGCGGGCAAGTCTACGTG 2239

Db 1477 GAGTCTCCGACGAGAGCACTGCGGAGCGCACACCCCAACAGCGCGGCGAGTCCACGAC 1536

Qy 2240 GGCACCAAGGCGCGG 2255

Db 1537 GCGTTGCGCGAGCGGG 1552

RESULT 5

US-09-103-840A-2

: Sequence 2, Application US/09103840A

: Patent No. 6294328

: GENERAL INFORMATION:

: APPLICANT: FLEISCHMAN, Robert D.

: APPLICANT: WHITE, Owen R.

: APPLICANT: FRASER, Claire M.

: APPLICANT: VENTER, John C.

: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

: TITLE OF INVENTION: TUBERCULOSIS

: FILE REFERENCE: 24366-2007.00

: CURRENT APPLICATION NUMBER: US/09/103,840A

: CURRENT FILING DATE: 1998-06-24

: NUMBER OF SEQ. ID NOS: 2

: SOFTWARE: PatentIn Ver. 2.1

: SEQ. ID NO 2

: LENGTH: 4403765

: TYPE: DNA

: ORGANISM: Mycobacterium tuberculosis

: FEATURE:

: OTHER INFORMATION: CDC 1551

: OTHER INFORMATION: "n" bases at various positions throughout the sequence

: OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match	2.6%	Score 71.4	DB 4	Length 4403765
Best Local Similarity	45.2%	Pred. No. 2.2e-06		
Matches	508	Conservative	0	Mismatches 598; Indels 18; Gaps 4
OY	1175	AACAATTATTCAGATGTACAGACGACGAGTACACCGATTCGATGGATCTCATCAAGCGCTCGGC	1234	
Db	3941794	AGCAGGGGTGGCGCGCGGACACCAACGGCTTCGGGGCGCCCGCGGCACCGCGGCAACAGC	3941853	
OY	1235	TGATGATTGAGTCTCTCGAGATTGACCCAAACGACGAACAACCTGGCTTAAGGGCAC	1294	
Db	3941854	GCGCCCGGGGGTGGTCATACGTTCAAGCAGCGCAGCAACGGCGGACCGCGGGCGCAC	3941913	
OY	1295	GGAAATGACAATCTTTGGGGGCGCAGATTCACCAACTGGGACACCGCGCACAAATGTGCA	1354	
Db	3941914	GGGGGCTGGCGGCGACCGGGGGGAGCGGGCGCAACGACGACCGCGCGCGGAGACCC	3941973	
OY	1355	ATCCAAATCACTGGCGAGACGGCATCGAGAAATTTCTCCGTCAAGACCTTGGCTTTCACCC	1414	
Db	3941974	GGCAAAAGCGGCGACCGCGCGGCGACCGGGCGCAACCGCGGGCGCGCGGTAGC	3942033	
OY	1415	GCGCGAAGCGAGCTATTGGCGCGCACTCGAGAGACGAACAAGGCTTCACCTTTTGCAAGCA	1474	
Db	3942034	GCGCGGGCGCAACTTCAACGCGCGGACCGCGGCGCACCGCGGACCGCGGCGGCGCGG	3942093	
OY	1475	AACGACCTCGGGAGACATCGCCCGAGACGGGCTGAGGGAAGGCCACATGGGGCCACCTCGACG	1534	
Db	3942094	AAAG-----GGGCGCTTAAACACCGACCGGACTCAAGACCGACCAACGAGCGACCGGCGC-	3942147	
OY	1535	AGCGTCGACTAGCGCGGAGACTCGGTCAAGAGACGTCTCTCCGTCGCGCAACACGCGCGC	1594	
Db	3942148	ACCGCGGCGCAACCGCGGCGCAAAAGCGGCGCACCGCGGGCGCGCGACGACTCCGCGGGGG	3942207	
OY	1555	ACGCAACAGTGGCCATCTCGTCGAGCGGCGCGCGGACGCTGGAGCATGACTACGCGGCC	1654	
Db	3942208	ACCGCGCGCAACGCGCGGGCGCGGCGGCGCAACGCGCGCGCGCGCTTAGCAACACCGCG	3942267	
OY	1655	GACACCTCATGAAAGGGGCGGACAGGTGGCCATTTCGGGCCAGCGGAGACAGATCTCTGG	1714	
Db	3942268	GCGACCGCGAGGCAACGCGGGGATCGGGGTGACGCGCGCAAGGCGGTAAACGGGGCGAA	3942327	
OY	1715	TGCAACCGCTCGTCGCGCGGCTGACAGGCTTCGAGATTTCAGAGGACGTTTGCTTCGTCG	1774	
Db	3942328	GGAAACAGCGGTTCCGAGATTGGGGG-----CGAGCCGGCTTTGGCGCGGGCCC	3942378	
OY	1775	AGCTGCCCGCGGGCGCGGTATCGCTTCGGAAGAAGACCAACAGCGCTTCTTACGCC	1834	
Db	3942379	GCGCGCAAAAGCGCGGGCGCGGCGGCGCACGCGGCGACCAACGGCTCGCGCGCC	3942438	
OY	1835	GGCTCCGGATCGACTTTTACGTACGCAAGAACACCGGACAGCTTACCGCGGGGCC	1894	
Db	3942439	GGCGGGCGCGGACAAAGCGGGCGCGGGGTCTCTGGCATCTTACCAACAGCGAGC	3942498	
OY	1895	AAGCTGGGCGAGCGGAGACGATCCGGGATATCGCTGCTCAACCCGACACCGCGGGGACG	1954	
Db	3942499	AAGCGCGGCAACCGCGGCGCACCGGGGCGCTGGGCGGCAACCGGGCGCACGGCGCAACGCA	3942558	
OY	1955	TTGTATGTCTGACCGAGCTGGCATTTTCGCTTCACAGACTCGGGCAGCAACCTTTGGC	2014	
Db	3942559	GCGACCGCGCGCGGCAACCGCGCAAAAGGGGCGACCGGGCGGACCGCGGCAACCGGGCG	3942618	
OY	2015	CAAGTCTCACCGCCCTTACCAACAACCTTACAGATGCGCCCTTGGGTGGGCTCAAGCTCG	2074	
Db	3942619	AGCGCGCGGGCGCGGTAGCGGGGCGCAACTTCAACGGCGCACCGGC---GCGAAC	3942675	
OY	2075	AACGTGAACCTGTATGCTTGGGACCGGGCCGTCGATAGGGGCTCGGCTTCAACGCACTGGA	2134	
Db	3942676	GCGCGCAAGCGCGGACCGCGGCGACAGGGGGCGCGGGCGCAACGGCGGCGCGGCCACCC	3942735	
OY	2135	GACACGCGCGCTTCGTGACGAGCATTCAGGGCTTCCAGGGCTTGGCTTCATCAACAGC	2194	
Db	3942736	GCGAGCGGCAACCGCGCAAAAGGGGCGCGCGGTGACCGCGGCAATGGCGCGCAAGCA	3942795	
OY	2195	ACCAAGTGGCGGCGGCGGACGACCGCGCGGCAAGTGTACTGTTGGGCAACCAAGCGCGG	2254	

Db 3942796 GGGCGACCCACCGGCGCTGGCGGAGCGGGCGGCGAGCGTGCTGTAACGGTGTGAACGGTGTGAAC 3942855

QY 2255 GCGCGTCTTTTACGCTCAGGAGAACCGTCGCGCGCGGCGGACGCGCGG 2298

||||| |

Db 3942856 GCGCGGACGCGGCGCTGCGCGCGGCGGCGCTGCGGCGGCGGACAG 3942899

||||| |

RESULT 6

US-08-209-747-1/c

Sequence 1, Application US/08209747

Patent No. 5733771

GENERAL INFORMATION:

APPLICANT: Lewis, Randolph V.

APPLICANT: Colgin, Mark

TITLE OF INVENTION: CDNAS Encoding Minor Ampullate Spider

TITLE OF INVENTION: Silk Proteins

NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22040-3487

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/209,747

FILING DATE: 14-MAR-1994

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1447-104P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-205-8000

TELEFAX: 703-205-8050

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2793 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Nephila clavipes

TISSUE TYPE: minor ampullate gland

FEATURE:

NAME/KEY: CDS

LOCATION: 183..2675

OTHER INFORMATION: /product= "N. clavipes minor

OTHER INFORMATION: ampullate silk protein"

US-08-209-747-1

Query Match 2.6%; Score 70; DB 1; Length 2793;

Best Local Similarity 41.0%; Pred. No. 3.4e-07;

Matches 325; Conservative 0; Mismatches 467; Indels 0; Gaps 0;

QY 1524 CCACTGACGAGGGGTGCACTACGCGGCGGAACCTGGTCAAGAGCGTGTCCGGCGGCA 1583

||||| |

Db 1222 CCACTTATACCAACGGTACCTTCACAGCTCTCTCTCTGACCTGCAGCGGCTCCAGCTCCA 1163

||||| |

QY 1584 ACACCGCGCGACGACGAACAGGTGGCCATCTGCTCGACGCGGCGCGACGTCGAGCATCG 1643

||||| |

Db 1162 GACACGACGACTCTACGTAACACACACACACTCCAGCTCTCGACACGACGACGACGCC 1103

||||| |

QY 1644 ACTACGCGGCGGACACGTCACATGAACGGCGGACACGGTGGCCTATTTCGGCCGACGGCGACA 1703

||||| |


```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/025,691
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/20120
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2712 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-025-691-4

```

Query Match 2.5%; Score 66.8; DB 3; Length 2712;

Best Local Similarity 44.5%; Pred. No. 1.9e-06;

```

Matches 405; Conservative 0; Mismatches 492; Indels 13; Gaps 3;

```

1515 CCACATGGCCACCTCGACGAGCTGACTAGCGCGGAACTGGTCAAGAGCGTCTCC 1574
 1598 CCATCGGTGACAGCTTGCCGCCCTACGACAGGCGTACACCGCGCGAGTGGTCCAGC 1757
 1575 GCGTCGGCAACAGCGCGGCGGCAAGAGTGGCCATCTGTCGAGAGGGCGCGGAGCT 1634
 1758 GCGTCGGCAACAGCTGGGACAGCGCGCTGCGGCAACTTCAACCAAGCTCCAGCTCA 1817
 1635 GAGCAGTCGACTAGCGCGGCGGACAGTCCATGAAGCGGCGACGCTGATTCGGCCG 1694
 1818 AGGCCAAGTACCGGACATCAAGTCTCTGCTTGGCGGCGGTGACCTGTGCCGCG 1877
 1695 AGCGGCAACAGATCTCTGCTGACACCGCTGTGCGCGCTGACCGCTCGAGCTTCCAG 1734
 1878 GCTTCACGAGCGCGGTGAAGAACCGCGCTTGGCCAACTCTGCACAGCACTGGTGC 1937
 1755 GCACCTTTGCTCCCTCTCGAGCTGCGCGCGGCGCGCTGATCGCTCGGAGAAAGA 1814
 1938 AGGACCGCGCGTGGCGGAGCTCTTCAGCGGCAATCGACTGAGTGAAGTACCGGAAG 1997
 1815 CCAACAGCGCTTCTAGCGCGGCTCCGAGTCACTTTTACGTACAGCAAGACACCGCA 1874
 1998 CCTCGCGCTCAGCTGCGAGACAGTCCGGTCCGGCGGCTGAAGAACATGCTCCAGGCG 2057
 1875 GCACCTTCAGCGCGGCGGCGGCAAGCTGGGCAAGCGAGGAGACATCCGGATATCCGCTC 1934
 2058 TGCGCGCCCACTTGGGACCGACCTGGTCAACCGCGCCATCACCGCGCAGCTCCG 2117
 1935 ACCGAGCAGCGCGGCGGCGGCTGATGCTCGACGAGCTGGGATATTCCGCTCCAG 1994
 2118 CGGCGAAGCTGAGCGCGCGGCGGAGTACG---CGGGCGCGCGCCAGTACTTTCAGTGTACA 2174
 1995 ACTCGGCGAGCACTTTGGCCAACTTTCACCGCGCTGACCAACACCTTACCAAGATCGCC 2054
 2175 ACGTGAATGACGTACGACTTTCTTGGCGCTGGGACAAAGACCGCGCGCAGCCGCCCACT 2234
 2055 TGGGTGTGGCTCAAGCTCGAACTGGAACCTGTATGCTTGGGACACGCGCGCGCTCAGGG 2114
 2235 CGGCGCTGAATCTCTACAGGCGGATCCCAAGGCGGAGTTCACATCGGCGCGCGCCATCG 2294
 2115 CTCGCTCTTACGCGAGTGAAGACAGCGCGGCTCTCTGAGCGAGATCCAGGAGCTCCAGG 2174
 2295 CCAAGCTTAAGGCGAAGGGGCTCCCGGAGAGCAAGCTCTCTGGGCTGCGCTTTCTACG 2354
 2175 GCTTGGCTTCATGAGACCAAGAGTCCGCGGAGCGGAGAGACCGCGCGGAGAG--- 2231
 2355 GCGCGGCTGAGACGGGCTTACCCAGGAGCGCGCGGCGGAGACCGCGCGGCGCA 2414
 2232 -----TCTAGCTGGGACCAAGCGCGCGGCGCTTTTACGCTCAGGGAACCGTCCGCG 2285

```

Db 2415 CCGGCACTACGAGCGCGGATCGAGACTACAGAGTCTCAAGAACCTGCCCCCA 2474
Oy 2286 CGGCGACGGCGGCGGACTTCTCTGACACCAAGACAGACAGACTCTTCCGCCA 2345
Db 2475 CCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2354
Oy 2346 GCTTCGACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2404
Db 2355 CCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2394
Oy 2405 TCGTCGAGCA 2414
Db 2395 TCTTCTGGGA 2604

```

RESULT 10

US-09-023-173-4

Sequence 4, Application US/09023173

Patent No. 6066781

GENERAL INFORMATION:

APPLICANT: Sutcliffe, Thomas D.

INVENTOR: Rodriguez, Raymond L.

TITLE OF INVENTION: Production of Mature Proteins

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

STREET: 350 Cambridge Ave., Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: IBM Compatible

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/023,173

FILING DATE: 13-FEB-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/038,168

FILING DATE: 13-FEB-1997

ATTORNEY/AGENT INFORMATION:

NAME: Petlihorly, Joanne R

REGISTRATION NUMBER: P42995

REFERENCE/DOCKET NUMBER: 0665-0007.30

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-324-0880

TELEFAX: 650-324-0960

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1140 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: codon-optimized Ramy3d signal-prosubtilisin BPN'

US-09-023-173-4

Query Match 2.4%; Score 65.4; DB 3; Length 1140;

Best Local Similarity 45.2%; Pred. No. 2.9e-06;

```

Matches 323; Conservative 0; Mismatches 386; Indels 6; Gaps 2;

```

1534 GAGCGTGCATACGCGCGGAACTGGGTCAAGAGGCTGTGCGGTGCGGCAACCGCGCG 1593
 312 GAGGTCCTCCGACGCGCTCAACAGATCAAGGCGCGCGCTTCCACAGCGGCGTACAC 371
 1594 CAGCAAGAGTGGCACTCTGCTCGAGCGGCGGCGGAGCTGAGCAATGACGCGGCG 1653
 372 CGGAGCAAGCTCAAGTCCCGCTCATCGACAGCGGCGGATGACAGCAAGCCAGCGGACT 431

```

OY 1654 CGACAGCTCATGAACGGCGGACGGTGGCTATTTCGGCCGACGGGACACGATCTCTG 1713
    || || || || || || || || || || || || || || || || || || || || ||
DB 432 CAGAGTCGGCGGGGGAGCTAGCATGCTGCCGAGCGGACCAACCGCTTCAGAGACCA 491
    || || || || || || || || || || || || || || || || || || || || ||
OY 1714 GTCGACCGCTCTGCTCCGGCTGCAGCGCTCGCATGTTCCAGGGCAGCTTTCCTCTC 1773
    || || || || || || || || || || || || || || || || || || || || ||
DB 492 CAGCCATGACCCACCAAGTCCGGCGGACCGCTCCCGCCCTCACCACAGCATGCGCTCT 551
    || || || || || || || || || || || || || || || || || || || || ||
OY 1774 GAGCCGCGCGGGGGGGCGGTATGCCCTCGGACAGAGAACAGACAGAGCTTTCACGC 1833
    || || || || || || || || || || || || || || || || || || || || ||
DB 552 CGGCGTCGCGCGGAGCGCTTACGCTCAGCCGCTCAAGTACTCGGCGGACGCGGCGG 611
    || || || || || || || || || || || || || || || || || || || || ||
OY 1834 CGGCTTCGGATGACCTTTTACGTAGCAAGGACACCGGACAGCTTCACGCGCGGCG 1893
    || || || || || || || || || || || || || || || || || || || || ||
DB 612 CCAGTACACTGGATGATCAAGGGGATGATGAGGGGATCGCCACACATGATGACGTCT 671
    || || || || || || || || || || || || || || || || || || || || ||
OY 1894 CAACTGGGACGAGGACGATCCGGGATATCGCTGCTCACCGACACCGCGGGGAC 1953
    || || || || || || || || || || || || || || || || || || || || ||
DB 672 CACCATGACCTTCGGGCGGCGCA--GCGGACAGCGCGCTCMAAGCGCGCTCGACAA 728
    || || || || || || || || || || || || || || || || || || || || ||
OY 1954 GTTGTATGCTGACCGAGCTGGGCAATTTCCGCTCCAGACTGGGGACGACCTTTGG 2013
    || || || || || || || || || || || || || || || || || || || || ||
DB 729 GGGCGTCGACAGCGGCTGTGCTGCTGCGCGCGCGGACAGAGGGGACAGCGGCGAG 788
    || || || || || || || || || || || || || || || || || || || || ||
OY 2014 CCAAGTCTCCACCGCCCTGACCAACACTACAGATGCGCTGGGTGTGGGCTCAGGCTC 2073
    || || || || || || || || || || || || || || || || || || || || ||
DB 789 CAGCAGCAGCTGCTGCTACCGCGGCAAGTACCGGACGTCATCGCGCTCGCGCGCTGGA 848
    || || || || || || || || || || || || || || || || || || || || ||
OY 2074 GAAGTGGAACTGTATGCTTGGGACCGGCGCGTCAAGGGGCTCGCTTACGCCAGTGG 2133
    || || || || || || || || || || || || || || || || || || || || ||
DB 849 CAGCAGCAACGACGCGGAGCTTTCAGCAGC--GTGCGCGCGGACCTGACGCTCATGGC 905
    || || || || || || || || || || || || || || || || || || || || ||
OY 2134 AGACAGCGGCGGCTCTCGGACGATCCAGGAGCTCCAGGGGCTTGGGCTCATGACAG 2193
    || || || || || || || || || || || || || || || || || || || || ||
DB 906 CCGGCGCTGACGATCAGAGACACCTCCCGGCAACAGTATGGGCGCTACAGCGGCGAC 965
    || || || || || || || || || || || || || || || || || || || || ||
OY 2194 CACCAAGTTCGCGGCGGACGAGCAGCAGCGCGGCAAGTCTACGTGGGACCAAC 2248
    || || || || || || || || || || || || || || || || || || || || ||
DB 966 CAGCATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
    || || || || || || || || || || || || || || || || || || || || ||

RESULT 11
US-09-408-647A-1
; Sequence 1, Application US/09408647A
; Patent No. 6399858
; GENERAL INFORMATION:
; APPLICANT: Kobayashi, Donald
; TITLE OF INVENTION: Chitinase Gene from Stenotrophomonas
; FILE REFERENCE: Rut-Cook 98-0090
; CURRENT APPLICATION NUMBER: US/09/408,647A
; CURRENT FILING DATE: 1999-08-26
; PRIOR APPLICATION NUMBER: 60/098,036
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2810
; TYPE: DNA
; ORGANISM: Stenotrophomonas maltophilia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (306)...(2405)
; OTHER INFORMATION: Open reading frame 1 (ORF1)
; NAME/KEY: sig_peptide
; LOCATION: (306)...(428)
; NAME/KEY: terminator
; LOCATION: (2457)...(2481)
; NAME/KEY: RBS
; LOCATION: (295)...(298)
; NAME/KEY: CDS
; LOCATION: (314)...(2162)

```

```

; OTHER INFORMATION: Open reading frame 2 (ORF11)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Genbank No. 6399858 AF014950
; DATABASE ENTRY DATE: 1997-09-23
US-09-408-647A-1

Query Match          2.2%; Score 60.2; DB 4; Length 2810;
Best Local Similarity 44.7%; Pred. No. 6,6e-05;
Matches 276; Conservative 0; Mismatches 338; Indels 3; Gaps 1;

OY 1777 CCTGCCCGCGGCGCGCTGATGCTCGCTGAGCAAGAACCAACAGGCTTCTTACGCGG 1836
    || || || || || || || || || || || || || || || || || || || || ||
DB 635 CCCACCGCGCGGCGGCGGCTTACCGCCGCGGACACATACCTGATAGGCCAACCGCGG 694
    || || || || || || || || || || || || || || || || || || || || ||
OY 1837 CTCGGATGACCTTTTACGTAGCAAGGACACCGGACAGCTTACCGCGGCGGCGCAA 1896
    || || || || || || || || || || || || || || || || || || || || ||
DB 695 CGACAGCGAGCGGACGTGTAGCAAGGTGAGTCTTCTCGTGCGGTACCTCGGTGGGCA 754
    || || || || || || || || || || || || || || || || || || || || ||
OY 1897 GCTGGGACGCGGACGAGGACATCCGGATATCGCTGCTCACCGGACCGCGGCGGACG 1956
    || || || || || || || || || || || || || || || || || || || || ||
DB 755 CGACACACGCGCGCGCTACAGCGGTACCTGGGCAATGATGATCGGCGGCGGACACCT 814
    || || || || || || || || || || || || || || || || || || || || ||
OY 1957 GTATGCTGACCGAGAGTGGCATATTCGGCTTCACAGACTGGGGACGACCTTTGGCCA 2016
    || || || || || || || || || || || || || || || || || || || || ||
DB 815 CAAAGGCGTGGCGCACGACCAACAAACGCGGTCACTTCTCGGCGGCGGCTCAGCGTGA 874
    || || || || || || || || || || || || || || || || || || || || ||
OY 2017 AGTCTCCACCGCGCTGACCAACACTACAGATGCGCTGGGTGTGGGCTCAGGCTCGAA 2076
    || || || || || || || || || || || || || || || || || || || || ||
DB 875 CGTCACTGCTCTCAGACACGACACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 934
    || || || || || || || || || || || || || || || || || || || || ||
OY 2077 CTGGAACCTGTATGCTTGGGACCGGCGCGTCAAGGGGCTCGCTTACGCCAGTGGAGA 2136
    || || || || || || || || || || || || || || || || || || || || ||
DB 935 CTCCAAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 994
    || || || || || || || || || || || || || || || || || || || || ||
OY 2137 CAGGCGGCGCTCTGAGCAGCATCCAGGCTCCCGAGGC---TTGGGCTCATGACAG 2193
    || || || || || || || || || || || || || || || || || || || || ||
DB 995 CAGTGGCTGGGCGGCTGACGCTGACGCTGACGCGGCGGCGGCGGCGGCGGCGGCGG 1054
    || || || || || || || || || || || || || || || || || || || || ||
OY 2194 CACCAAGTTCGCGGCGGACGAGCAGCAGCGCGGCAAGTCTACGTGGGACCAACGGCGG 2253
    || || || || || || || || || || || || || || || || || || || || ||
DB 1055 CACCGATGACGAGGAGGCGGCGGCTCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1114
    || || || || || || || || || || || || || || || || || || || || ||
OY 2254 GGGCGTCTTTACGCTCAGGGAACCGTCCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2313
    || || || || || || || || || || || || || || || || || || || || ||
DB 1115 CCGCGACACAGCGCGGCAATGCTTCGCGGCGGAGAGCGGCTGATCAGGTACACCGCGG 1174
    || || || || || || || || || || || || || || || || || || || || ||
OY 2314 CAAAGAGAGAGGACGAGTACCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2373
    || || || || || || || || || || || || || || || || || || || || ||
DB 1175 CGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1234
    || || || || || || || || || || || || || || || || || || || || ||
OY 2374 TGTATCCAGGACCGCGG 2390
    || || || || || || || || || || || || || || || || || || || || ||
DB 1235 CGGCGGCGGCGGCGGCGG 1251
    || || || || || || || || || || || || || || || || || || || || ||

RESULT 12
US-09-103-840A-1/C
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 441529

```


OY	368	CCGAGTATGAGGCAATCATCTTCCTCGTCAAGACCGGGGCGCAAGTGGCTTTCACCAAC	427
Db	1494	-CCGGCACTCGCCAGATGTTGGCCTCTGTTCAACCGGGCGCACTGTTGAGAGCTTCCGAC	1552
OY	428	TTGGCCCTTCAAAAGTCGGGGGTAACATGCCAGACGGGAGCCGGAAGAGCTTGGCTGTC	487
Db	1553	CTGGCGTTTAAGCTGGGTGGTAACCAAGCTGGCCGGCCCAATGGGAGGGGCTGGCGGTG	1612
OY	488	GATCCGGCCAACTCCAACTATCTACTTGGTGTCTGCTCAGAGAAAGGCTCTGGAG	547
Db	1613	GACCCGACGATGGCCCGCGTGTCTGTGGGCTCCGG---GATGGCGGCTGTGGCGT	1669
OY	548	TCCTACGACGCGCGCTGACCTTTTCCAAAGTCTCTGCTTCACGCAACTGGAGCTAC	607
Db	1670	AGGCACGATCGCGGGCGGCACCTGGCGGAAGTGGCTCTTTCGGAGCCGCGCTGGCC	1729
OY	608	ATCCCAAGACCCGAGTATGTTCCAAAGGCTACACAGCGCAAGCAAGACTATGGGTT	667
Db	1730	GGTGCCACACCGGCGCCATTCATGTTGGGCGCGA---GCAAGCGGTGGGATCCCTTTGTC	1786
OY	668	ACGTTGCACTCAACACGACAGACGACCGGGGAGCCAGCTCTGATCTTTGTGGCAG	727
Db	1787	GTTGTCCAGCAGGCACTGGCAACACAGGCTGGCCAAACCCGCAATCTACGTGGGCGTG	1846
OY	728	GCTGATACATCACTGCTTTCAGTCTTATGTAGCAGCAATAGCCGCGCTCCACGTGAGTCT	787
Db	1847	TC---CACCGAAGACACGACGACTGTATGTGTCCGAAGATGCGGGCGGAGTTGGGCACCG	1903
OY	788	GTACCGGGGACGACGAGGAATACTTTCTCTACAGGGCAAACTGCAGGCCACAGAGAG	847
Db	1904	GTGGCCCGGGCAACACCCGCGCTGTGCGCC---GAGCCACATGGCGCGGCGCACCGATG	1958
OY	848	G---GCTTGTATCTGACCTATTCGATGGAGGACAGGCGCGTATGATGGACACTTGGCTAG	905
Db	1959	GGCACTGATCTGAGCTTATGGCGACAGCCGCGGCGGACCTGATGGCCGGGGAGCCT	2018
OY	906	TGTGAGAGTACGACATTTGACGGGGGAATTTGGAAAGACATCAACCCCTGTCTCTGGATCAG	965
Db	2019	TGTGGAAATTACAGCCCGGACAGAGGGGCGGTGGGTGATACGCCGATTCGCGAGCCAG	2078
OY	966	ATCTATATCTTTGGCTTTTGGCGGCGCTTGGCGCTGATTTGCAAAAGCCA	1012
Db	2079	CCAGTGGCGAGTATGCTGGGTGGGTGGTGGCGGTGATCCGAA	2125
US-09-748-033-6			
US-09-748-033-6			
Sequence 6, Application US/09748033			
Patent No. US20020069431A1			
GENERAL INFORMATION:			
APPLICANT: Broadway, Roxanne M.			
TITLE OF INVENTION: EFFECT OF ENDOCHITINASE AND CHITOBIOSTIDASE AND THEIR			
FILE REFERENCE: 19603/3091			
CURRENT APPLICATION NUMBER: US/09/748,033			
PRIORITY FILING DATE: 2000-12-22			
PRIORITY FILING DATE: 1999-12-23			
NUMBER OF SEQ ID NOS: 8			
SOFTWARE: PatentIn Ver. 2.1			
SEQ ID NO 6			
LENGTH: 1107			
TYPE: DNA			
ORGANISM: Streptomyces albidoflavus			
US-09-748-033-6			
Query Match			
Best Local Similarity 4.5%: Score 66.8; DB 10; Length 1107;			
Matches 405; Conservative 0; Mismatches 492; Indels 13; Gaps			
1515 CCACATGGGCGACCTGCAGCAGCGCTGACATACGCCGGAAGTGGGTCAAGAGCTGTCC 1574			

D	131	CCATCGGTGACAGTTTGCCCGCTTAGACAAAGGCGTACACCGCCGCGAGTGGCGACG	190
Q	1575	GCGTGGCAACACCGCCGGCAGCAACAGGTGGCCATCTCGTCAGACGGCGCGGACGT	1634
D	191	GGCGTCGCGGACACTGGGAGCACCCCGCTCGGGGCAACTTCAMCAGGTCCGCAAGCTCA	250
Q	1635	GGACATCGACTACGGGGCCGACAGTCCATGAAAGCGGGCAGGGTGGCTATTGCGGCG	1694
D	251	AGGCCAAGTACCCGGCACATATCAAGTCCCTTGCTTCCTGGCGGGCTGGACCTGTGCCG	310
Q	1695	ACGGGACACAGATCTCTGGTTCACCGCCCTGTCGGGCTGCAAGCGCTCGAGTTCACG	1754
D	311	GCTTACACGAGCGCGGTGAAGAACCCGGCGCCTTGGCCAAATCTGACACGACCTGGTCG	370
Q	1755	GCAGCTTGGCTCCGTCTGAGAGCTGACCCGGGGGGCGGTATCGGCTCGGACAAAGA	1814
D	371	AGGACCCGCGCTGGCGGAGCTTTCGACGGGATCGACTCGACGTGGAGTAACCGAAGC	430
Q	1815	CCAACACGCTTCTTACGCGCGGCTCCGATTCGACCTTTACCTGCAAGGACACGGCA	1874
D	431	CTGTGGGCTCTACGTGTGAGAGCTCCGGTCCGGCGCGGTGAAGAACATGGTGCACGGGA	490
Q	1875	GCACCTTTCACGCGCGGGCCCAAGCTGGGACAGCGGCAAGGACGATCCGGATATCGCTGCTC	1934
D	491	TGCGCGCCCATGTTTCGGGACCGACCTGTGCACGCGCGCCATCCGCCACGCGCACGTCCG	550
Q	1935	ACCCGACACACCGGGGGAGCTGTATGTCTCGACGAGTGGGCAATTTCGGCTCCACAG	1994
D	551	CGCGCAAGCTCGACGCCGCGGACTAG---CGGGCGCGCCCAAGTACTTTCGACTGTACA	607
Q	1995	ACTGGGACAGACTTTTGGCCAAATGTCACGCGCCCTGACCAACACTACAGATCGCC	2054
D	608	ACGTGATGACGTACGACTTCTTGGGCGCTGGGAGAACACGGCCGACCGGCCACT	667
Q	2055	TGGGTGTGGGCTCAGGCTCGAAGTGAACCTGTATGTCTTCGGGACCGGCCCTGACGGG	2114
D	668	CGGCGCTGAAGTCTCTACAGGGGCGATCCCAAGGCGCACTTCACATCGGCCCGGCATCG	727
Q	2115	CTCCGCTCTACGGCAGTGGAGACAGGGGGCCCTCCGTGACGGACATTCACAGGCTCCAGG	2174
D	728	CCAAGCTCAAGGGGAAGGGGCTCCGGGACGACAGTCTCTGCTCGGCAATCGGCTTTACG	787
Q	2175	GCTTGGCTCATGACAGACCAAGTGCGGGACGCGGACGAGCAGCAGCGCGGGCAAG---	2231
D	788	GCGCGGCTGAGACGGCGGTACCCACGAGGAGCCCGGGCGGACCGCCACCGCGCGGCA	847
Q	2232	-----TCTACGTGGGACCAACAGGCGGGGGGCTCTTTACGCTCAAGGAACCGTCCGGC	2285
D	848	CGGCGACTACGAGGGGGGATGAGGACTACAGGTTCTTCAAGAACCTGCCCCGGCA	907
Q	2286	GGGCGACGGGCGGAGCTTCTCTGTCACCAAGCAGAGCAGCAGCATCTTCCTCCGCA	2345
D	908	CGGCGACCGTGGGGGACCGGCTTACGCCAATGGGGGACGCAACTGTGTGAGACTACGCA	967
Q	2346	GCTTGACACCAAGCTGAGGTGA--GCCTTGTATCCAGACCCGGGCTTTCGACGCTACT	2404
D	968	CCCCGGCACATCAAGACCAAGATGACCTGGGCGCAAGGACCGGGGCTTCGGCGCGGCT	1027
Q	2405	TCGTGAGGA	2414
D	1028	TCCTTGGGA	1037

RESULT 3
 US-09-748-033-4
 : Sequence 4, Application US/09748033
 : Patent No. US20020069431A1
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Broadway, Roxanne M.
 : APPLICANT: Gongora, Carmenza E.
 : TITLE OF INVENTION: EFFECT OF ENOCHITTINASE AND CHITOBIOSIDASE AND THEIR
 : TITLE OF INVENTION: ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT


```

Db 699 GAAGTCTCTGCCAAGCCGGAAGCAAGCCGCGGACGACGCGGTTTCTCGGCTTC 758
Qy 2149 CTGAGCAGACATCCAGGCTTCCAGGCTTCCATCCATGACAGCAAGTCCGCGG 2208
| 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 759 CATGACACACAGATGGGCAATCCAGGCAAGCCGACCTGGTGTATGAACTTCGAGCGCG 818
Qy 2209 CAGCGGACAGACCGCGGCGGAGTCTACGTGGGACCAACGCGCGGCGCTT 2262
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 819 CACGCGCTCTGTATCGGCGCGCGGACGAGGCGCTGTAACTGATGTTCACCTT 872

RESULT 7
US-09-864-761-19241/c
; Sequence 19241, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19241
; LENGTH: 1075
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL078472.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1

```

```

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
; OTHER INFORMATION: EXPRESSED IN PLACENT, SIGNAL = 20
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
; OTHER INFORMATION: EST_HUMAN HIT: AV739739.1, EVALUE 1.00e+00
; OTHER INFORMATION: NT HIT: AL163201.2, EVALUE 2.00e-19
US-09-864-761-19241

Query Match 2.28; Score 59.2; DB 10; Length 1075;
Best Local Similarity 43.08; Pred. No. 2.4e-07;
Matches 289; Conservative 0; Mismatches 383; Indels 0; Gaps 0;

Qy 1765 CTCGCTCTGAGCTTCCCGGCGGCGGCTATCGCTCGGACAAAGAACCAACAGCT 1824
| 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 772 CACCACACACACCATCACCACATACCATCATCATCAGCAACCACTACACACCATCAC 713
Qy 1825 CTTTACGCGCGGCTCGGATCGACCTTTTACGTGAGCAAGAGACCGGACAGCTTAC 1884
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 712 CACCACACACATCACCACACACATCACCACACACACACATCACCACATCAC 653
Qy 1885 GCGGCGGCGGCGGAGCGGAGGAGAGATTCGGGATATGCTGCTCACCACCGACAC 1944
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 652 CACTACACACACACACACACACATCACCACATCACCACATCACACACATCACAC 593
Qy 1945 GCGGCGGAGCTGTATGTCTGACCGAGAGTGGCATATTCGCTCCACAGACTCGGCGAC 2004
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 592 CACCACCATCACCACATCACCACACACACACACACATCACCACACATCACCACAC 533
Qy 2005 GACCTTTGGGCAAGTCTCCACCGGCTGAGCAACCTTACAGATGCGCCCTGGGTGGG 2064
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 532 CACCACACACATCACCACATTAACACAGACACACACACACACACACATCACCACAT 473
Qy 2065 CTCAGGCTCGAATGGAACCTGTATGCTGCGGACCGGCGGCGGCTGCGCTCTA 2124
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 472 CACTACCATCTACACACACACACACATCACCACATCACCACACACACATCACAT 413
Qy 2125 CGCGAGTGGAGACAGCGGCGCTCTGAGAGAGATCAGAGGCTCCAGGCTTCGGCTC 2184
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 412 CACCAGACACACACATCACCACACACACACAGTACACATCACCACATCACCACAC 353
Qy 2185 CATGACAGACACCAAGTTCGCGGAGGAGAGACACCGCGGCGGCAAGTCTACGTGGGAC 2244
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 352 CATCACCACACACATCACCACATCACCACATCACCACACACACATCACCACATAC 293
Qy 2245 CAACGCGCGGCGGCTCTTTTACGCTCAGGGAACGCTGGGCGGCGGCGGAGCTTC 2304
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 292 CATCACCACATCACCACACACACATCACCACATCACCACACACACATCACCACAC 233
Qy 2305 CTCGTGACCAAGCAGAGACAGACAGTACCTTTCGCGCAGCTGAGACACAGCTGAG 2364
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 232 CACCATCACCACATCACCACACACACACATCACCACATCACCACATCACCACAT 173
Qy 2365 GTGAGCGTTGTATCCAGACAGCCGCGGCTTGAGCGGTGATCTGTCGAGGACACTGGC 2424
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 172 CACCACACACATCACCACACACACATCACCACACACACACACACATCACCACAC 113
Qy 2425 CGCGGTCGCCAC 2436
| 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 112 TACCACATCACAC 101

RESULT 8
US-09-864-761-2513/c
; Sequence 2513, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

```



```

Oy 1790 GCCGTCATGCGCTCGGACAGAGAACACAGCGCTTCTACCCCGGCTCCGATGAC 1849
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 64 GCGGCGCTTGCGCGAGTTATCTCCACCGTCATCTGTGCTTGCGCGGACAGCGCTCG 123
Oy 1850 TTTTACGTACGACAGACACCGGACGAGCTTACGCGCGGCGGCGCAAGCTGGCAGCG 1909
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 124 ATGGCGCTTACGAGAACTGACCGCGGCGAGCGACCGCGCGCGCGGATCGGCGG 183
Oy 1910 GGGACATCCGGGATATGCTGCTCAACCGGACCGCGGCGAGCTTGTATGTCTGAC 1969
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 184 GCGGTGCGGCGCGCTTCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 243
Oy 1970 GACGTCGGATATTCGCTCCACAGACTCGGCGACGACCTTTGGCCAGTCTCCAGCG 2029
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 244 GCGCGATGAGAACCGCGCGGACCTTCCGCGGCTTCTGCTGCGGCGGCAACATCAG 303
Oy 2030 CTGACCAACACCTACAGATCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 2089
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 304 GCGGCGCTCTCTACGATCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 363
Oy 2090 GCGTTCGCGACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2149
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 364 GCGTTCGCGACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 423
Oy 2150 TGGACGAGATCCAGGCTCCGCGGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCT 2209
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 424 GAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 483
Oy 2210 AGCG-----GCGACCGCGCGGCGAGCTTACGTGGGACCAAGCGCGGCGCTTT 2263
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 484 GCGGTTGACCGCGAGAGGCGCGCTGCGGACCATGCGCGCGCGCGCTTCTGCTGCT 543
Oy 2264 TAGGCTCAGAGGAAGCTCGCGGCGGCGGCGGAGCTTCTGCTGCTGCTGCTGCTGCT 2323
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 544 GTCGCGCGGCAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
Oy 2324 AGCAGCATGCTCTTCCGCGGCGCTG 2350
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 604 TCGTTCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 630

RESULT 10
US-09-887-576-784
; Sequence 784, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; PRIOR FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 784
; LENGTH: 1185
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-887-576-784
Query Match 2.2% Score 58.4; DB 10; Length 1185;

```

```

Best Local Similarity 43.8%; Pred. No. 4,3e-07;
Matches 351; Conservative 0; Mismatches 446; Indels 5; Gaps 2;

Oy 1567 CGTGTGCGCGCTGCGGACACCGCGGCGACGCAAGGTGGGATGCTGCTGCGAGCGG 1626
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 168 GCGGAGATACACCAAGGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 227
Oy 1627 GCGGAGTGGAGATGAGTACGCGGCGGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 1686
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 228 GCGCATCGGCTTGTGCTGCGAGCTGCGGCGCTGCGAGCGGCGAGCGCTGCAAGGCTGCT 287
Oy 1687 TTGCGCGGCGGCGGACGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1746
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 288 CAACATGAGGACGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 347
Oy 1747 GTTCGAGGCGAGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1806
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 348 CCGCGAGGAGATCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 407
Oy 1807 CAGAAGACCAACAGCTGCTTACGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1866
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 408 CCGCGAGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 467
Oy 1867 CACCGGAGAGCTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1926
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 468 GGTTCGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 524
Oy 1927 GCGTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1986
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 525 GGTGTGATGCTTACAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 584
Oy 1987 CTCCACAGACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2046
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 585 CTCCACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 644
Oy 2047 GATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2106
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 645 GGTGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 704
Oy 2107 GTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2166
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 705 CCGCTGCGGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 764
Oy 2167 CTCCAGGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2226
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 765 GATCATCATGACCATGAGGCGGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGAA 824
Oy 2227 GCAAGTCTAGTGGGACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2286
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 825 GGAACCGGACCAAGGTGACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 884
Oy 2287 GCGGACGCGGCGGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2346
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 885 GCTGCGGCGGCGG--CCTGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 942
Oy 2347 CTGAGCACCACGCTGAGGCTG 2368
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 943 CCGGAGCGGCTTCCGCTGCTG 964
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

RESULT 11
US-09-894-844-60/c
; Sequence 60, Application US/09894844
; Patent No. US20020176873A1
; GENERAL INFORMATION:
; APPLICANT: Behr, Marcel
; APPLICANT: Small, Peter
; APPLICANT: Schoolnik, Gary
; APPLICANT: Wilson, Michael A.
; TITLE OF INVENTION: Molecular Differences Between Species of
; FILE REFERENCE: STANI02CON
; CURRENT APPLICATION NUMBER: US/09/894,844

```

```
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/318,191
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/097,936
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Mycobacteria tuberculosis
US-09-844-60
```

```
Query Match      2.1%; Score 58; DB 9; Length 1173;
Best Local Similarity 44.8%; Pred. No. 5,7e-07;
Matches 223; Conservative 0; Mismatches 275; Indels 0; Gaps 0;
```

```
OY 1763 GCCTCCGCTGAGACCTGCCCCGGGGCGCGTATGCTCGGACAGAAAGACACAGC 1822
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 704 GCCTGCCCCACCTGACAGCGCGCATCCACCGGCAACCCCGCCGACCTGCGCGCA 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1823 GTCCTTACGCGCGCTCGGATGACCTTTTACGTACAGAACAGACCGGAGAGCTTC 1882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 644 GCCACACACAGCGGACCCCTCCACGACACCTCCGCGATCGGAGAACAGCGACGAA 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1883 ACGCGCGGCGCAAGCTGGGACGAGGACGATCCGCGATATCGCTCACCGCGAC 1942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 584 ACGGACGCTGCGCATCCCGGTCACTGCGACCGCGCTGGGAGGCAACCCCGCAATTC 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1943 ACGCGGCGACGTTGTATGTCTGACCGACGCTGGCATATTCGCTCCACAGACTCGGGC 2002
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 524 ACGCGCGGACACTTAAAGCGGCTGACGCTCACCGCGCGCGCGCGCTGTATAC 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2003 ACGACCTTTGGCGAAGTCTCACCGCGCTGACCAACACTACAGATCGCTGGGTG 2062
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 464 CCCACCTTCGACACCACTGCTGAGCCACATCTTCACATATGAACTGACGCGCA 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2063 GCGTACGCTGAACTGAACTGATGCTTCGCGACCGCGCGCTGAGGGGCTCGGCTC 2122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 404 ATGCGCGCGGTCTGACCCAAATGTGTGCGCAACAGCGCCCAACACACCGGA 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2123 TAGCCGAGTGAGACAGCGCGCTCTGAGACGACATCCAGGGCTCCAGGGCTTGCGC 2182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 344 TTCCCTCTACCGCGCGCGGTGACCGTGGCCGCAAGCGCGCTCAAGCGCTGCGC 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2183 TCCATCAGACAGACAGTGGCGGAGCGAGACCGCGCGGGGCAAGTCTACGTGGGC 2242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 284 GCGCGGCTAGGCTTACGACGCGGACACTCGGCTGCGCGCGCGCGCACTAACACGCC 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2243 ACGAAGCGCGGGGCGTC 2260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 224 ACATACGCGCGCGCGCGC 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 12
US-09-860-846-7
; Sequence 7, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding mythemycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1248
```

```
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-7
```

```
Query Match      2.1%; Score 56; DB 9; Length 1248;
Best Local Similarity 43.9%; Pred. No. 2,4e-06;
Matches 338; Conservative 0; Mismatches 425; Indels 7; Gaps 2;
```

```
OY 1663 CATGACGCGCGGACGCTGAGGCTTATGCGGCGAGCGGACAGCATCTGTTGACCGC 1722
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84 CATGACGCGCGGACGCTGATGACGCGGCTGAGCGGCGCTCGACAGCATGCTGCTC 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1723 CTCGTCGCGGCTGACGCGCTGCTGATTCAGGGGACGCTTTCCTGCTGAGCTGCC 1782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 144 CAAGGCGGCGCGCTGCTGCTGCGGATTCGAGGAGCGGCTGCGCGGCTGCGCGGCTCG 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1783 GCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 204 GCATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1843 ATGACCTTTTACGTACAGAACAGACCGGACGCTTACGCGCGGCGCGCAAGCTGG 1902
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 264 CTTGACCGCGGACGATGATGATGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTG 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1903 CAGCGAGGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1962
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 324 CTGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1963 CTGACGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2022
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 384 GACACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2023 CAGCGCGCTGACCAACACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2082
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 444 GCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2083 CTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 501 GCTGTACTTGCACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2143 GCGCTCTGAGAGCATCTGAGGCTTCCAGGGCTTGGCTGCTGCTGCTGCTGCTGCT 2202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 561 CTTGCGGACCGCGAGGCTTTCACCTTCCAGCCACCAAGGCGCTTTCAGGAGG 620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2203 GCGCGAGCGGACAGACCGCGCGGCAAGTCTAGTGGGCAACAGCGCGCGGCTGCT 2262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 621 GCGGCGCGTCTGACAGGAGCGGACGCTGCGCGCGGATCCGCGCTTCCACACTT 680
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2263 TTACGCTCAGGAGCGCTGCGCGGCGGCGGCGGCGGCACTTCTGCTGACCAAGCAG 2322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 681 GCGCTTGCAGCTGCGCGGCGGACCGCGCGGCGGCGGCGGCAAGGATGAGCGAG 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2323 CAGCAGC---AGTACCTTCTCCGCACTGAGCAGCAGCAGCTGAGCTGAGCTTGTAT 2378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 741 GCGCGCGCGCATGCGGCTTCCAGCTTCCAGGCTGACGAGCGGAAACCG 800
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2379 CCAGACCGCGGCTTTCAGAGGCTGCTGAGAGCAGCTGCGCGCGC 2428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 801 GCGCAACGCGCGCGCTTACGCGGAGCAGCAGCTGCGGAGCTTCCCGCGCTC 850
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 13
US-09-988-384B-7
; Sequence 7, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding mythemycin and pikromycin
; FILE REFERENCE: 600,536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
```

```
: CURRENT FILING DATE: 2001-11-19
: PRIOR APPLICATION NUMBER: PCT/US99/14398
: PRIOR FILING DATE: 1999-06-25
: PRIOR APPLICATION NUMBER: US 09/105,537
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 53
: SEQ ID NO 7
: LENGTH: 1248
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-988-384B-7

Query Match      2.1%: Score 56; DB 9; Length 1248;
Best Local Similarity 43.9%: Pred. No. 2.4e-06;
Matches 338; Conservative 0; Mismatches 425; Indels 7; Gaps 2;

OY 1663 CATGACGCGCGCAGCGGTGCTATTCGCGGACGCGGACACGATCTGTGTCGACCGC 1722
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84 CATGACGCGCGCAGCGGTGCTATTCGCGGACGCGGCTGACCGGCGCTCGACGAGTGTCTC 143

OY 1723 CTCGTCGCGCGGTGCGAGCGCTGCGAGTTCGAGGCGAGCTTTGCGTCCGCTGAGCGTCCG 1782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 144 CACGCGCGCGCGCGCTGCTGCGGAGTTCGAGAGCGCGCTGCGCGGCTGCGCGGCTCG 203

OY 1783 GCGCGCGCGCGCTGATCGCTCGGACAGAGACCAACAGCGCTTCTTACGCGCGCTCGG 1842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 204 GCATCGCGCTGCGGACCTGCAACGCGGCGCGGCTCGAGCTCCCTCGGCGACGCGCGCGG 263

OY 1843 ATCGACCTTTTACGTCACAGAGGACACCGGCGAGCTTTCACGCGCGGCGCGGCGGCG 1902
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 264 CCTGACCGCGGAGATGATATGCGCTGATGACGTTTCGCGCGCGACCGCGACGACTCG 323

OY 1903 CAGCGCAGGAGAGATCGCGGATATGCTGCTACCGCGACACCGCGGCGAGCTGTATGT 1962
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 324 CTGAGTTCGCGCTGACACCGCGGCTTTCGCGGACATCGACCGCGGACACCGGCACTCGACCC 383

OY 1963 CTGACCGCAGCTCGGCAATATTCGCTCCACAGACTCGGCGACGACTTTCGCGCAAGTCTC 2022
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 384 GGACGAGGTGCGCGCGCGGCTGACACCGCGACCTCGGCGCGCTGCTGCGGCGCTGCTG 443

OY 2023 CACGCGCGCGGACCAACACTTACAGATGCGCGCTGCGGCTGAGGCTGCAACTGGA 2082
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 444 GGGCGCGCGCTGCGCGCGCGCGGACGACTG---CGGAGGTCGCGGACGAGCAGCGGCTCGG 500

OY 2083 CCTGTATGCTCTTTCGCGACCGCGCGCTGACGAGGCTTCGCTTACGCGCAGTGAAGACGCG 2142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 501 GCTGTACTTTCAGACGCGCGGCGGACGCGCTGCGGCTGCGGCTGCGGCGCGCGCGGCGAG 560

OY 2143 GCGCTCTGAGAGGACATTCAGAGGCTCCAGGCGCTTCGCGCTCATGACAGCAGCAAGGT 2202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 561 CCTCGCGCAGCGCGAGGCTTTCAGCTTCACGCGCACAAAGCGCTTCACAGCGCTTCGAGGG 620

OY 2203 GCGCGGCGGACGCGGACACCGCGCGGCGGCAAGTCAAGTGGGCGACCAAGCGCGGCGCTT 2262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 621 CCGCGCGCGCTGCTGACGAGAGCGGCGGACTTCGCGCGCGGATTCGCGGCTTCACAACTT 680

OY 2263 TTACGCTGAGGGAACCTGCGGCGGCGGCGGCGGAGCTTCTGCTGACCAAGCAGAG 2322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 681 CCGGCTTTCAGACTGCGCGGCGGCGGAGCGCGCGCGGCGGAGCAAGGCAAGATGAGGAGCG 740

OY 2323 CAGCAGC-----AGTACTCTTTCGCGCAGCTGAGCAACCAAGCTGAGGTGTTAT 2378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 741 CCGCGCGCGCATGAGGCTTCACCTCTCGACGCGCTTCCTCGAGGTTCATGACGCGCAACCG 800

OY 2379 CCAGCAGCGCGGCTTCGAGGCTGAGTGTGAGAGCAAGCGTTCGCGCGCGGCTTC 2428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 801 CGGCAACACGCGCGCTTCACGCGGAGCAGCTTCGCGGAGCTTCGCGCGGCTTC 850
```

RESULT 14
US-09-861-289-7
Sequence 7, Application US/09861289
Patent No. US20020110897A1

```
: GENERAL INFORMATION:
: APPLICANT: Sherman, D.H.
: APPLICANT: Liu, H.
: APPLICANT: Xue, Y.
: APPLICANT: Zhao, L.
: TITLE OF INVENTION: DNA encoding methymycin and pikromycin
: FILE REFERENCE: 600,438US1
: CURRENT APPLICATION NUMBER: US/09/861,289
: CURRENT FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 09/105,537
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 7
: LENGTH: 1248
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-861-289-7

Query Match      2.1%: Score 56; DB 10; Length 1248;
Best Local Similarity 43.9%: Pred. No. 2.4e-06;
Matches 338; Conservative 0; Mismatches 425; Indels 7; Gaps 2;

OY 1663 CATGACGCGCGCAGCGGTGCTATTCGCGGACGCGGACACGATCTGTGTCGACCGC 1722
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84 CATGACGCGCGCAGCGGTGCTATTCGCGGACGCGGCTGACCGGCGCTCGACGAGTGTCTC 143

OY 1723 CTCGTCGCGCGGTGCGAGCGCTGCGAGTTCGAGGCGAGCTTTGCGTCCGCTGAGCGTCCG 1782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 144 CACGCGCGCGCGCGCTGCTGCGGAGTTCGAGAGCGCGCTGCGCGGCTGCGCGGCTCG 203

OY 1783 GCGCGCGCGCGCTGATCGCTCGGACAGAGACCAACAGCGCTTCTTACGCGCGCTCGG 1842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 204 GCATCGCGCTGCGGACCTGCAACGCGGCGCGGCTCGAGCTCCCTCGGCGACGCGCGCGG 263

OY 1843 ATCGACCTTTTACGTCACAGAGGACACCGGCGAGCTTTCACGCGCGGCGCGGCGGCG 1902
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 264 CCTGACCGCGGAGATGATATGCGCTGATGACGTTTCGCGCGCGACCGCGACGACTCG 323

OY 1903 CAGCGCAGGAGAGATCGCGGATATGCTGCTACCGCGACACCGCGGCGAGCTGTATGT 1962
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 324 CTGAGTTCGCGCTGACACCGCGGCTTTCGCGGACATCGACCGCGGACACCGGCACTCGACCC 383

OY 1963 CTGACCGCAGCTCGGCAATATTCGCTCCACAGACTCGGCGACGACTTTCGCGCAAGTCTC 2022
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 384 GGACGAGGTGCGCGCGCGGCTGACACCGCGACCTCGGCGCGCTGCTGCGGCGCTGCTG 443

OY 2023 CACGCGCGCGGACCAACACTTACAGATGCGCGCTGCGGCTGAGGCTGCAACTGGA 2082
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 444 GGGCGCGCGCTGCGCGCGCGCGGACGACTG---CGGAGGTCGCGGACGAGCAGCGGCTCGG 500

OY 2083 CCTGTATGCTCTTTCGCGACCGCGCGCTGACGAGGCTTCGCTTACGCGCAGTGAAGACGCG 2142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 501 GCTGTACTTTCAGACGCGCGGCGGACGCGCTGCGGCTGCGGCTGCGGCGCGCGCGGCGAG 560

OY 2143 GCGCTCTGAGAGGACATTCAGAGGCTCCAGGCGCTTCGCGCTCATGACAGCAGCAAGGT 2202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 561 CCTCGCGCAGCGCGAGGCTTTCAGCTTCACGCGCACAAAGCGCTTCACAGCGCTTCGAGGG 620

OY 2203 GCGCGGCGGACGCGGACACCGCGCGGCGGCAAGTCAAGTGGGCGACCAAGCGCGGCGCTT 2262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 621 CCGCGCGCGCTGCTGACGAGAGCGGCGGACTTCGCGCGCGGATTCGCGGCTTCACAACTT 680

OY 2263 TTACGCTGAGGGAACCTGCGGCGGCGGCGGCGGAGCTTCTGCTGACCAAGCAGAG 2322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 681 CCGGCTTTCAGACTGCGCGGCGGCGGAGCGCGCGCGGCGGAGCAAGGCAAGATGAGGAGCG 740

OY 2323 CAGCAGC-----AGTACTCTTTCGCGCAGCTGAGCAACCAAGCTGAGGTGTTAT 2378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 741 CCGCGCGCGCATGAGGCTTCACCTCTCGACGCGCTTCCTCGAGGTTCATGACGCGCAACCG 800

OY 2379 CCAGCAGCGCGGCTTCGAGGCTGAGTGTGAGAGCAAGCGTTCGCGCGCGGCTTC 2428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 801 GCGCAACCGCCGCTACCGGAGCAGCTCGCGGACCTCCCGGCGTCC 850

RESULT 15

US-09-988-384B-3/C

Sequence 3, Application US/09988384B

Publication No. US20030073824A1

GENERAL INFORMATION:

APPLICANT: Sherman, D.H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600,536US1

CURRENT APPLICATION NUMBER: US/09/988,384B

CURRENT FILING DATE: 2001-11-19

PRIOR APPLICATION NUMBER: PCT/US99/14398

PRIOR FILING DATE: 1999-06-25

PRIOR APPLICATION NUMBER: US 09/105,537

PRIOR FILING DATE: 1998-06-26

NUMBER OF SEQ ID NOS: 53

SEQ ID NO 3

LENGTH: 12441

TYPE: DNA

ORGANISM: Streptomyces venezuelae

US-09-988-384B-3

Query Match 2.1%; Score 56; DB 9; Length 12441;

Best Local Similarity 43.9%; Pred. No. 7, 2e-06;

Matches 338; Conservative 0; Mismatches 425; Indels 7; Gaps 2;

QY 1663 CATTGAACGGCGGACGGTGGCTATTTCGCGGAGCGGAGACGATCTTGTGACCGC 1722

Db 12113 CATCGACCGCGGCGCTGTACGAGCGGCTCGACCGGGCCCTCGACAGCGGTGCTC 12054

QY 1723 CTGTCGGGCGGTGACGGCTCGAGGTTCAGGGGAGCTTGGCTCCCTCGAGCTGCC 1782

Db 12053 CAACGGGGGCGGCTGCTCGCGGAGGAGCGGCTCGCGGGCTCGCGGGGCTCGG 11994

QY 1783 CCGGGCGCGGCTGATCGCTCGGAGAGAGACACAGCGCTTCTTACGGCGGCTCGG 1842

Db 11993 GCATGCGGCTGGCCACCTGACAGCGGCGGCGGCTCGAGCTCTCGCGACCGCGCGG 11934

QY 1843 ATGACCTTTTACGTACGAGGAGACCGGAGCGGCTTACCGCGGCGGCGGCGGCGG 1902

Db 11933 CCTACCGGCGGAGTGTATGCGGTGATGAGTTCGCGCGCACCGCGGAGCATGCGG 11874

QY 1903 CAGCGGAGGAGCATCCGGATATCGCTGCTCACCGGAGCGGCGGCGGAGCTTGTATGT 1962

Db 11873 CTGGATGCGGCTGACCGGCTTTCGCGGAGATGACCGGAGACCGGCAACCTCGACCC 11814

QY 1963 CTGACCGAGCGTGGCATATTTCGCTCCACAGACTCGGGGACGACTTGGCCAACTCTC 2022

Db 11813 GAGACAGGTGGCGGCGGCGGCTCACACCGGCGGCTCGGCGGCTCGGCGGCTCGG 11754

QY 2023 CACCGCGGCTGACCAACACTACAGATCGGCTGGGTGTGGGCTCAGGCTGAACTGAA 2082

Db 11753 GGGCGGCGGCTGCGGCGGCGGAGCTG---CGGAAAGTTCGCGGAGGAGGCGGCTGCGG 11697

QY 2083 CCGTATGCTTGGGCGGCGGCGGCTGAGGGGCTCGGCTTACGCCAGTGAAGACAGCGG 2142

Db 11696 GGTGTACTTGGAGCGGCGGCGGCGGCTGCGGCTGCGGCGGCGGCGGCGGCGGCGG 11637

QY 2143 CCGCTCTGAGCGGAGATCCAGGAGGCTCCAGGCGCTTGGCTCCATCGACAGCAAGTCTC 2202

Db 11636 CCGTGGGAGAGCGGAGGCTTCAAGCTTCAAGCGGCGGCTTCAAGCGGCTTCAAGCGG 11577

QY 2203 CCGCGGAGCGGAGAGACCGCGGCGGCAAGTCTACGTGGGAGCAACGCGGCGGCGGCTT 2262

Db 11576 CCGCGGCGGCTGACAGGAGCGGCGGAGCTTCCGCGGCGGAGTCCGCTTCCAACTT 11517

QY 2263 TTACGCTCAAGGAGCGGCTCGGCGGCGGAGGCGGCGGAGCTTCCGCTTCCAAAGCAGAG 2322

Db 11516 CCGCTTGGACCTGCGCGGCGGAGAGCGGCGGCGGCGGAGCAACGCCAAGATGAGCGAGCC 11457

QY 2323 CAGCAGC---AGTACCTTTCGCGGAGCTGAGACACAGCGTGAAGGTGAGGCTGTAT 2378

Db 11456 CCGCGGCGGCGGAGGCTTCACTTCCCTGAGAGCGGCTTTCGCGAGGTATCGAGCGGAGCCG 11397

QY 2379 CCAGGACCGGCGGCTTTCGCGGAGCTTTCGAGAGACCAAGCTCGGCGGCGG 2428

Db 11396 GCGCAACCGCGGCGGCTTACCGGAGGAGCACTGCGGAGCTCCCGGCGGCTCC 11347

Search completed: May 4, 2003, 06:50:35
Job time: 248.979 secs

THIS PAGE BLANK (USPTO)

LOCUS	BO751758	426 bp	mRNA	linear	EST 18-JUL-2002
DEFINITION	EST632321 DSCT Colletotrichum trifolii cDNA clone pDSCT7-37, mRNA sequence.				
ACCESSION	BO751758				
VERSION	BO751758.1 GI:21907163				
KEYWORDS	EST.				
SOURCE	Colletotrichum trifolii.				
ORGANISM	Colletotrichum trifolii.				
REFERENCE	Eukaryote; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae; mitosporic Phyllachoraceae; Colletotrichum.				
AUTHORS	1 (bases 1 to 426) Samac,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T., Cheng,F. and Fraser,C.M.				
TITLE	ESTs from mycelia of Colletotrichum trifolii race 1				
JOURNAL	Unpublished (2002)				
COMMENT	Other_ESTs: EST632322 Contact: Deborah A. Samac Department of Plant Pathology University of Minnesota 495 Borling Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA Tel: 612 625 1243 Fax: 651 649 5058 Email: debbys@puccini.crl.umn.edu r1GR sequence name: MTSAG377K More information is available at: www.medicago.org Seq primer: SKmod (CTA GAA CTA gta gat CC).				
FEATURES	Seq primer: SKmod (CTA GAA CTA gta gat CC).				
SOURCE	Location/Qualifiers				
	1..426				
	/organism="Colletotrichum trifolii"				
	/strain="race 1"				
	/db_xref="taxon:5466"				
	/clone="pDSCT7-37"				
	/clone_lib="DSCT"				
	/tissue_type="mycelia"				
	/dev_stage="young, actively growing mycelia (3 days after inoculation) grown in liquid culture (cutin minimal medium containing 2%glucose)."				
	/lab_host="DH5alpha"				
	/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: EcoRI; Isolate: 2sp2; cDNA was prepared from polyA+ enriched RNA. The cDNA was ligated into lambda gtl1 from Stratagene and packaged using Gigapack packaging extracts. An aliquot of the amplified library was used to transduce E. coli Y1090 and phage DNA was purified from a liquid lysate. The cDNA inserts were gel purified after EcoRI digestion and ligated into pBluescript SK+. Aliquots of the ligation were used to transform E. coli DH5alpha which were plated onto medium with X-gal for selection of recombinants."				
BASE COUNT	77 a 152 c 106 g 91 t				
ORIGIN					
Query Match	4.1%; Score 111.4; DB 14; Length 426;				
Best Local Similarity	70.8%; Pred. NO. 1.6e-14;				
Matches 148; Conservative	0; Mismatches 61; Indels 0; Gaps 0;				
0Y	53	CGAGTCTTGCCCTTGCTGCTGGGGGCGCTCATCCTGCCCATGCTGCTTTTCATGGAAG	112		
Db	186	CGAGGGGTCGGGGCTCTCTCGGCGGCGCCCGCTCTCGTCCGGCGGCTGCGACTGGAAA	245		
0Y	113	AACGTCAAGCTCGGGGGGGGGGGGCTTCGTCGCCCGGCATCATCTTCCATGCCAAGACA	172		
Db	246	AACGTGCACACGGGGGGGGGGGGGGGGGGCTTCGTCGCCCGGCATCATCTTCCATGCCAAGACA	305		
0Y	173	AAAGCGTAGCATATAGCAGAACAGATATTGGCGGGCTGATACCGGCTCAACGCCGACGAC	232		
Db	306	AAGGCGTCGCGCTTGCGCCCGAACAAGATATCGGGGGCGCTGTACGGCTCAACGCCGACGAT	365		
0Y	233	TGATGACACCGCCGCTACGAGATGGATTGC	261		
Db	366	TCTTGACTTCGCTACCGACGCCCAACGC	394		

[illegible]

Df	627	GBSNSSGGGSGSVASAGMSGVSSVSSSCGRSSSOGCGGCVCGSGSSSSGSGSGGSVSVC	568
Oy	2451	CCGG 2454 : :	
Df	567	CSSG 564	
<hr/>			
RESULT 5			
BE0A1072/c			
LOCUS	BE0A1072	699 bp	mRNA linear EST 07-JUN-2000
DEFINITION	OF19D05 OF Oryza sativa cDNA 5' similar to transcription factor,		
ACCESSION	BE0A1072		
VERSION	BE0A1072.1	GI:8336782	
KEYWORDS	EST.		
SOURCE	Oryza sativa.		
ORGANISM	Oryza sativa		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
AUTHORS	Ehretiolidae; OrYZaeae; Oryza. 1 (bases 1 to 699)		
TITLE	Bonnett,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferreira, H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacios,C., Scara,G., Wheeler,M., and Zepeda,G.R. Functional Genomics of Plant Stress Tolerance Unpublished (2000) Contact: Michalowski,C.B. University of Arizona Bio Sciences West room 513, Tucson, AZ 85721, USA Tel: 520-621-7982 Fax: 520-621-1697 Email: cbm@u.arizona.edu		
JOURNAL	An open reading frame exists. location/qualifiers		
COMMENT	1..699 /organism="Oryza sativa" /strain="Pokkali" /db_xref="taxon:4530" /clone_1lb="OF" /tissue_type="entire plant" /dev_stage="2 weeks" /note="1 week 150mm NaCl"		
<hr/>			
BASE COUNT	158 A 242 C 189 G 109 T	1 others	
ORIGIN			
<hr/>			
Query Match	2.4% Score 64.6 DB 10 Length 699		
Best Local Similarity	53.1% Pred. No. 0.00054		
Matches	161 Conservative 0 Mismatches 139 Indels 3 Gaps 1		
OY	1540 CGACTACGCCGGGAAGCTGCTCAGAAGAGGCGTCCGCTCGGCAACACC GCCGACGCA	1599	
Df	510 GGATTCGTGCGTGCGAGATTCCCACAACCGGAGCGGCGGCGCATTCGGCGCGC	451	
OY	1600 ACAGTTGCCATTCTGTCGCCAGCGGGCGGCGCACGTGAGCATCGACTACGGGCGGACAC	1659	
Df	450 GTAGAGCTGCTGATCGGCTTCCGCCGCGCGCATGTGCTTCAGAGCTGTCATGTTTA	391	
OY	1660 GTCATGAAAGCGGCGACAGGTGAGCTATTGGCCGACGCGACAGATCTTGTCGAC	1719	
Df	390 GTTCATGCTCAGCAAGCTGCTGCGGCGGCTGTGGCGCCCGCACACAG---CTGCTCAG	334	
OY	1720 GCCTCTGTCGGCGGTGACGCCCTGCGAAGTCCAGGCGACGTTTGCTTCGCTCGAGCT	1779	
Df	333 GGCCTCTGTCGCGCGCGCGACGTCGCTTTTGTGACGCGCGGCGATGCTGTCATGCT	274	
OY	1780 GCCCGCGGGGCGCGCATCGCTCGGAAAAGAACCAACAGCGTCTTACGCGGCTC	1839	
Df	273 CTGTCTCCCGGTGAGCAGAGAGAACAGAAACGAGAGGTCTGTGGCGCGGCGGA	214	
OY	1840 CGG 1842		
Df	213 TCG 211		

LOCUS	B1306418/c		445 bp	mRNA	linear	EST 20-JUL-2001
DEFINITION	NL_4_011 Drought stress (leaf) Oryza sativa cDNA clone NL_4_J11 3', mRNA sequence.					
ACCESSION	B1306418					
VERSION	B1306418.1					
KEYWORDS	EST.					
SOURCE	Oryza sativa.					
ORGANISM	Oryza sativa. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehretidoideae; Oryzaceae; Oryza. 1 (bases 1 to 445) Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabushan,I., Ravindrababu,P. and Bennetzen,J.L. Novel EST enrichment with normalized cDNA libraries from drought stressed rice (Oryza sativa L.cv Nagina 22) Unpublished (2001)					
TITLE	Contact: Reddy AR					
JOURNAL	Department of Plant Sciences, School of Life Sciences University of Hyderabad P.O. Central University, Hyderabad-500 046, A.P, India Tel: 0091-40-3010265 Fax: 0091-40-3010145 Email: arjuls@uohyd.ernet.in Insert length: 445 Std Error: 0.00 Plate: 4 row: 3 column: 11 Seq primer: GTAAACCAGCGCAGTC. Location/Qualifiers					
FEATURES	1..445					
source	/organism="Oryza sativa" /cultivar="Nagina 22 (Indica sub sp)" /db_xref="taxon:4530" /clone="NL_4_011" /clone_1lb="Drought stress (leaf)" /tissue_type="Entire leaf tissue" /dev_stage="35 day-old seedlings" /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized leaf cDNA library from drought stressed seedlings"					
BASE COUNT	86 a 172 c 134 g 53 t					
ORIGIN						
Query Match	2.3%; Score 62.4; DB 13; Length 445;					
Best Local Similarity	53.9%; Pred. No. 0.0016;					
Matches	151; Conservative 0; Mismatches 126; Indels 3; Gaps 1;					
OY	1553 AGAGCTGCTCCGGCTGGCGCAACACCGCGGCAGCACAGAGTGCGCATCTGTCCGACG 1622					
DB	445 AGACGCCGACGCGCGCGGCGGACATCCAGCGCGCGCTGTACACACTGTAATGCGCTTCG 386					
OY	1623 GCGCGCGCACTGAGCATTCAGCTAAGCGGGCGCGAACAGTCATGAAGCGCGGCACGTTGG 1682					
DB	385 CGCGCGCATATCGTTCAGAGCTGTGTCATAGTGTATCATCTCATGACAGCTGGGCTCG 326					
OY	1683 CCTATTCCGCCGACGCGGACGATCTCTCTGTGGTCAGACCGCTGTCGCGCGTAGAGCGCT 1742					
DB	325 CCGGCTGTGTGGCGCGCGCCACACAG--CTGTGTAGAGGCGCTGTGTGTGGCGCGGACG 269					
OY	1743 CGCAGTTCAGAGGACACTTTCCTCCCTGCTGTGAGCTTCGCCGCGGCGCGCTCATGCGCT 1802					
DB	268 TGCTGCTTTCTTGTGAGCGCCCGACGCGCGCTGCTGTGTGTGCGCCGTGTGAGCGAGG 209					
OY	1803 CGGACAAGAAGACCAACAGCGCTTCTAAGCGCGGCTCGG 1842					
DB	208 AGGACAAGAACCGAGGAGGTCTGTGCGCGCGGCGATGG 169					
RESULT 7						
LOCUS	B1305544/c		458 bp	mRNA	linear	EST 20-JUL-2001
DEFINITION	NL_0_103 Drought stress (leaf) Oryza sativa cDNA clone NL_0_103 3',					

[illegible]

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	Genoscope.	Direct Submission		
2	Submitted (02-JUN-1999)	Genoscope - Centre National de Sequencage :		
3	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr			
4	- Web : www.genoscope.cns.fr)			
5	Determination of this BAC-end sequence was carried out as part of a			
6	collaboration with the Berkeley Drosophila Genome Project (BDGP).			
7	The BDGP is constructing a physical map of the Drosophila			
8	melanogaster genome using these BACs. For further information			
9	please see http://www.fruitfly.org The BDGP Drosophila			
10	melanogaster BAC library was prepared by Kazutoyo Osoegawa and			
11	Aaron Mamoser in Pletier de Jong's laboratory in the Department of			
12	Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,			
13	NY. The library is named RPc1-98 and was constructed by partial			
14	EscoRI digestion of Drosophila DNA provided by the BDGP from the			
15	isogenic strain y2; cn bw sp, the same strain used for the BDGP's			
16	p1 and EstR libraries. A more detailed description of the library			
17	and how to order individual BAC clones, the entire library, or			
18	filters for hybridization from the BACPAC Resource Center can be			
19	found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .			
20	Location/Qualifiers			
21	1..925			
22	/organism="Drosophila melanogaster"			
23	/db_xref="taxon:7227"			
24	/clone="BACR19D16"			
25	/clone_lib="RPc1-98"			
26	/note="end : TET3"			
27	BASE COUNT	120 a	61 c	61 g
28	ORIGIN		172 t	511 others
29	Query Match	2.2%	Score 60.8:	DB 17; Length 925;
30	Best Local Similarity	14.0%	Pred. No. 0.0039;	
31	Matches	52; Conservative 173;	Mismatches 147;	Indels 0; Gaps 0;
32	OY	1544	TACGCGGGGAATCGTCGATCAAGAGCGTCGCGCGGCAACACCGCGGACGCAACAG	1603
33	DB	553	TSSSSSTGKGGKSSSSGBSCSCSSSCSSSCCBCCSCSSCYCCSSBSBSKSS	612
34	OY	1604	GTGGCCATCTCGTCGACGGCGCGCGACGTCGATGACATGACTACGCGCCGACACGTC	1663
35	DB	613	TBSGCCSCCKSKSVCTCCSSSSSSSSSSSTSSSTSSSTSSKSSSGSSSSSYTTTS	672
36	OY	1664	ATGAACGGCGGACGCTGATTCGCGCGGACGCGGACGACATCTCTGTGTCACCGCC	1723
37	DB	673	KSTASGSGSMWAGGSGSTGTSSTSSSSSSSTSSSSSVSGSKSTBSBSGSSSSGSS	732
38	OY	1724	TGCTCGCGGCGTCACCGCTCGCATGTCACGGGACACTTCCCTCGCTCGACGCTGCC	1783
39	DB	733	SSSTSSBBSCTSTSSSSSSSYSTTCSCCTCCSYSTSSSTSSSTSSSTSSGSSSV	792
40	OY	1784	CGGCGCGCGCTATGCGCTCGGACAAGAACAACACAGCTTCTTACGCGCGCTCCGA	1843
41	DB	793	GTSSSSDSTSTCCSCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	852
42	OY	1844	TGCACTTTTACGTACGACAGACACCGGACGACAGCTTTCACGCGCGGCGCCCAAGCTGG	1903
43	DB	853	TGCMGTSTACSSSSSSSSSSSSSVSSSSSSSVSSSSSVSSSSSVSSSSSVSSSSSV	912
44	OY	1904	AGCGGACGAGC 1915	
45	DB	913	SSGSGSGSGSVS 924	
46	RESULT 9			
47	LOCUS	B0463043	559 bp	mRNA linear EST 30-MAY-2002
48	DEFINITION	HI02N16r HI Hordeum vulgare cDNA clone HI02N16 5'-PRIME, mRNA		

sequence.
 accession B0463043
 version B0463043.1 GI:21270825
 keywords EST
 source Hordeum vulgare.
 organism Hordeum vulgare.
 Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Trilicaceae; Hordeum.
 1 (bases 1 to 559)
 reference Zhang, H., Weschke, W., Michalek, W., Stein, N. and Graner, A.
 authors EST sequencing and analysis in barley (2002)
 title Unpublished (2002)
 journal Contact: Stein Nils
 comment Molecular Markers Group, Department Genbank
 Institute of Plant Genetics and Crop Plant Research (IPK)
 Corrensstr. 3, 06466, Gatersleben, Germany
 Corrensstr. 3, 06466, Gatersleben, Germany
 Tel: 039482-5522
 Fax: 039482-5595
 Email: stein@ipk-gatersleben.de
 Insert length: 559 Std Error: 0.00
 Plate: 2 row: N column: 16
 Seq primer: M13rev.
 FEATURES
 source Location/Qualifiers
 1. 559
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone="H102N16"
 /clone_1lb="H1"
 /tissue_type="female inflorescences"
 /dev_stage="female inflorescences (approx. 3 mm in size)"
 /lab_host="XL10-Gold"
 /note="Vector: pBluescript SK-; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA). Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable."
 BASE COUNT 75 a 279 c 122 g 83 t
 ORIGIN
 Query Match 2.2%; Score 60.2; DB 14; Length 559;
 Best Local Similarity 46.5%; Pred. No. 0.0051;
 Matches 194; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

Db 500 AGCCCCCGCATGCGGACCGCCCTCCATCGACACACACCTGACCGTGTGGACATCG 556
 RESULT 10
 B0620053
 LOCUS 629 bp mRNA 11near EST 28-JUN-2002
 DEFINITION Talr1134H10F Talr1 Trilicium aestivum cDNA clone Talr1134H10F, mRNA
 sequence.
 accession B0620053
 version B0620053.1 GI:21625132
 keywords EST.
 source bread wheat.
 organism Trilicium aestivum
 Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Trilicaceae; Trilicium.
 1 (bases 1 to 629)
 reference Cloutier, S., Dong, G. and Walsh, A.
 authors Wheat functional genomics - Thatcher Lr1 cDNA library
 title Unpublished (2001)
 journal Contact: Dr. Sylvie Cloutier
 comment Cereal Research Centre, Agriculture and Agri-food Canada
 195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
 Tel: (204) 983-2340
 Fax: (204) 983-4604
 Email: scloutier@em.agr.ca
 was cloned directionally, not all sequences generated with reverse
 primer where from the 5' end (same with forward primer and 3' end).
 Average insert size is >2.2Kb
 Plate: 134 row: H column: 10
 Seq primer: M13 Forward.
 FEATURES
 source Location/Qualifiers
 1. 629
 /organism="Trilicium aestivum"
 /cultivar="Thatcher Lr1"
 /db_xref="taxon:4565"
 /clone="Talr1134H10F"
 /clone_1lb="Talr1"
 /tissue_type="leaf tissue"
 /dev_stage="14 Days old"
 /lab_host="E. coli XL0R"
 /note="Vector: Lambda ZapII; mass excised in plasmid
 vector pBK-CMV (Stratagene); Site 1: EcoRI; Site 2: XhoI;
 mRNA obtained from wheat NIL Thatcher Lr1 24 hours after
 inoculation with leaf rust pathogen Puccinia triticina
 race BB6 carrying the avirulence gene Avr1."
 BASE COUNT 88 a 233 c 196 g 112 t
 ORIGIN
 Query Match 2.2%; Score 59; DB 14; Length 629;
 Best Local Similarity 45.3%; Pred. No. 0.0096;
 Matches 215; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

BASE COUNT	82 a	202 c	233 g	92 t	2 others
ORIGIN					

Query Match	2.18;	Score 56.6;	DB 9;	Length 611;
Best Local Similarity	45.68;	Pred. No. 0.033;		
Matches 200;	Conservative 0;	Mismatches 239;	Indels 0;	Caps 0

QY	1467	CCGCAAAAGACCTCGGGACATTCGCCGAGACGGTCTGGGCAACGCCCAATGGGGCA	1526b
Db	454	CGACCCCGCGCGGTGATGACGACACACGGGGGTGAGTGTGGTTCACACACCTTGGGGCC	395.
QY	1527	CCCTGCAGCAGCGTTCGACTACGCCGGGACCTCGGTCAAGAGCGCTCGCGCTCGGCACCA	1586b
Db	394	CGCGCTCTGTGGGACACCGCGGCCAGCATGGGATCTCGGCGACGGGCAAGCGTGGGTGG	335
QY	1587	CGCGCGGACGACAAAGTGGCCATCTCGTCGAGCGGGCGGCGACGTTGGAACATCGACT	1646b
Db	334	ACATGCTCTCCACGATCACCAACCCGGGTCCTCCCGCGGCTCTATGGCCGCGGCGACAGCGCG	275
QY	1647	ACGGGGCGACAGTCCATCAAGGGGGGACAGGTGGCCATATCGGCGCGACGGCGACAGCA	1706b
Db	274	CCCTGCTCTCGCGGTCCACCAACGTCGCTCTACACGCCGAGGTGGCGCGGACAGAACCGGG	215

RESULT 13			
LOCUS	BE490392		
DEFINITION	BE490392	487 bp	mRNA
ACCESSION	WHE0367_H11_021s	wheat cold-stressed seedling cDNA library	linear
VERSION	WHE0367_H11_021	mRNA sequence.	
KEYWORDS	BE490392.1	GI:9610021	
SOURCE	EST.		
ORGANISM	bread wheat.		
	Triticum aestivum		

```

      . 85 a      186 c      147 g      69 t
BASE COUNT
ORIGIN
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHEU0367_H11_021"
/clone_id="Wheat cold-stressed seedling cDNA library"
/tissue_type="Seedling"
/dev_stage="Five-day old seedling"
/lab_host="E. coli SOLR"
/note="Vector: lambda uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
, germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered-crystallization dishes. Five-day
old seedlings were transferred to 5 C cold room and kept
for 48 hr. The tissue, total RNA, and poly(A) RNA were
prepared, a cDNA library was made, and the cDNA clones
were in vivo excised to give Bluescript phagemids in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
```

Query Match	2.1%	Score 56	DB 10	Length 487
Best Local Similarity	47.2%	Pred. NO. 0.045		
Matches 170; Conservative	0	Mismatches 190	Indels 0	Gaps 0

Accession	Sequence	Position
Oy	CTTACCTTTGCGAGCAAGACACCTCGGACATCTGGCCGACAGAGCTGTGGCAAGGCC	1515
Oy	1456 CTTACCTTTGCGAGCAAGACACCTCGGACATCTGGCCGACAGAGCTGTGGCAAGGCC	1515
Dh	14 CGTATATATGCGTGGCCGCATCTACTCAAGACCCGCGCGGCCACAGGCGGGCGGCAT	73
Oy	1516 CACATGGGCCACTCTGCAGAGCGGTGCATACCGCGGAACTCGTCTAAGAGCGTGTCCG	1575
Dh	74 GCCCAAGGGGAGCTACACCATCTGTGCACCTGGACACCTCGGCTCCCTCTGAAAGCGGTCCCA	133
Oy	1576 CGTGGGAAACCGCGCGGACGCAACAGGAGGACATCTGTCCAGAGGCGGGCGAGCGG	1635
Dh	134 GTTGGTCAAGAACGTGGCGCAGCTTCGACATGCCCATCGAGAGTGTGTGTGTGCAAGCGCCG	193
Oy	1636 GAGCATCGACTACGCGGCGGCACACAGTGCATGAACGGCGGCACAGTGGCTATTTCGGCGGA	1695
Dh	194 CGTATACAGGCCACCCGCAAGAGAGGCTTCTCTTACCGCGGACGGGCTTTCGAAATCAGGT	253
Oy	1696 CGGAGACACGATCTCTTGATGCGACCGCGCTGTCTCGGCGTGCAGCGCTGCAGATTTCAGG	1755
Dh	254 CGGCGTCAACCACTGTGGGCCATTTCTCTCTCGCGCGGACATCTCTGAGGACCTCAAGGC	313
Oy	1756 CAGCTTTGGCTCGGTCTGCAGGCTGCCCGCGGCGCCGTATGCGCTCGGACAGAAGAC	1815
Dh	314 CTTCGAGTACCCCTTCCAAGGCGCTCATCATCTGTGGCTCATACCGGGAACACCAACAC	373

REFERENCE	1 (bases 1 to 487)
AUTHORS	Anderson O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hata, C.C., Kang, Y., Lao, G.R., Miller, R., Rausch, C.D., Seaton, C.L. and Tong, J.C.
TITLE	The structure and function of the expressed portion of the wheat genomes - Cold-stressed seedling cDNA library
JOURNAL	Unpublished (2000)
COMMENT	Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel.: 5105395773 Fax: 5105595818 Email: oanderson@wp.usda.gov Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: Stragane SK primer.
FEATURES	Location/Qualifiers
SOURCE	1. 487 /organism="Triticum aestivum"

Wed May 7 14:15:34 2003

Search completed: May 4, 2003, 08:22:33
Job time : 2506.47 secs

us-10-026-994-1.rst

DR WPI; 2001-656860/75.
 DR N-PSDB; ABL03526.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 PS
 XX Disclosure; SEQ ID NO 5061; 21pp + sequence listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
 CC (AB057737-AB072072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX Sequence 1472 AA:
 SO
 Query Match 5.7%; Score 246.5; DB 22; Length 1472;
 Best Local Similarity 20.8%; Pred. No. 1.7e-07;
 Matches 203; Conservative 110; Mismatches 363; Indels 301; Gaps 41;
 51 TDGIADNAGMNMGIDVALDPODDOKVYAVGVMTNSWPDNSGALIRSSDR----- 102
 178 TRGSSDSRGWRGRTRENERMORRE-----PWSQONAGODHGDNDNRANDNRGQ 227
 103 -----GATWSFTNLFPKVGGMGPGAGERLAVDPANSNIIFGARSNG- 147
 228 RNGGGRGPGGGGGGVSRSRGGGGMGRGTGPRGDRSGPGAGYSGRGANED 287
 148 -----LKSITDGGVTFPSKVSFTATGTYIPDPSNGYNSKQGLMYTFPS-----TSST 198
 288 HHEVELMDNT-----IAONAEKQOAHDAWDAMWNEEYEGSL--KDSKVFTTSLN 336
 199 TGCATSRIFVGTADNITASYVSTNAGSTWASVPG-----QPKYFPKAKLQPAEKAL 252
 337 ATQSAANVSTGTGASVTA--VPAAGTEISAPGLEHQLVQSGSHLESSSSPPAAVTP 393
 253 YLT-----YSDGTGPGYDGLGVSRYDIAGTWDITP--V 286
 394 PATLGSAGATPPLQYSAVSNPPOLOSGTGAGTGSASA-----AAGCAGASTPSSFV 448
 287 SGLSLYGFEGGLG-----DIQKRGTLVVASLNSWMPDQLF----- 323
 449 SASPDTFSSAASAATLVHQAQKQOOLQOQTPPIKPSATLSEQSYFNSLASOGVSPGS 508
 324 ---RSTSGTWSPIWAM-----ASYPTETYYYSISIPKA-----PMIK 359
 509 VPQOASAGVANOAVAAVQSSTSVGSQYR-NTYANVFASGTAAGTAQSOQOQPIR 567
 360 NMFIDV-----TSESPDGLIKRLGWM--TESLEIDPTDSNHL--YGTGNTIF 404
 568 RARVKLPPEPSKIPASAVEMPCDNLNINIGYLDVQFGLDFTDGFELPEKVGSGSID 627
 405 GGHULTMWDTRHNVSIC-----SLADGIEFESVODLASAPG-----GSELLAA 447
 628 GQOQOQOQPDVQSKSQOQOQVTLAAGLOSSQISDALNAGTYSRSTSOQOQGVSAVNAT 687
 448 V-----GDDNGFTFASRNDLQTSPOTWATPT-----WATSTSV- 482
 688 IDOLTKRDPICQGTGSGNATONAVOSSGASKTASGPTTAPAGCIGSSSTYANVQSVANSY 747
 483 -----YAGNSVKSIVRVCTACTOYVAISSDGATWSIDYADTSMN--GGTVAYSA 531
 748 QOQGYGVSQSSVQOQAGSGAGSTGAVSGGCGTATQNIPIGSSSQNSTSGNMSAYLT 807
 532 DGGTILMSTASSGVQRSGQFQSGFASVSSLRAGAVIASDKTNSVFYAGSGSTFYVSGDTG 591
 532 DGGTILMSTASSGVQRSGQFQSGFASVSSLRAGAVIASDKTNSVFYAGSGSTFYVSGDTG 591

Db 808 SG-----YSTPOSAYOSSQ-----SYVGNWTGLSNSSGFGSGASN---ASSOY 846
 QY 592 SSFTRGPKLSAGTIRIDIAHPTTAGTLYVSTDVGIKRSIDSGTTPQOVAFALNTVOIA 651
 Db 847 ANFSASAKLKA-TTASAAHYDS-----VSTSGV--SSNSGST----- 883
 QY 652 LGVSGSNMNLVYARGTGPGSARLYASGASWPDIOGQGFSGIDSTKVAGSGSTAGOV 711
 Db 884 ---GNG-----GVVSGQTGANQAAVSNNS-----VSGSSSVSNV--TAGVASGNVAG-V 927
 QY 712 YVINGKRVFTYAGTGVGGTGTSSSTKQSSSTSSASSTLLASVYST----- 762
 Db 928 GGGVSGSGSVSGVGPDSASVGVNWNNSSSASVGAATVAGTATGATVAVLASLTNK 987
 QY 763 -----RASTYSSRSTSSAAGPTGCGVGHVAGQCGIGMTG-----PT--OC 801
 Db 988 NTSSSNSSGSGSAATTTGNNSGGAGASTGCV-GSSSGAGAGSGGSGGLVPTNIOM 1046
 QY 802 VAPVCCOKNDYVYQCV 818
 Db 1047 VSQYI-QTGLPYQQPV 1062
 RESULT 5
 AAY00865
 ID AAY00865 standard; Protein: 536 AA.
 XX
 AC AAY00865;
 XX
 DT 21-MAY-1999 (first entry)
 XX
 DE Cellulohydrolase CBH B protein sequence.
 XX
 KW Cellulohydrolase; CBH A: CBH B: cellulose degradation; food production;
 KW beverage production; animal feed production; endoglucanase.
 OS Aspergillus niger.
 OS
 PN WO9006574-A1.
 XX
 PD 11-FEB-1999.
 XX
 PF 31-JUL-1998; 98WO-EP05047.
 XX
 PR 31-JUL-1997; 97EP-0202389.
 XX
 PA (KONN) GIST-BROCADES BV.
 XX
 PI De Graaff LH, Gielkens MWC, Visser J;
 XX
 DR WPI; 1999-153802/13.
 DR N-PSDB; AAX22095.
 XX
 PT New cellulohydrolase, enzymes from Aspergillus - used to degrade
 PT cellulose, e.g. in food, beverage, animal feed, pulp, paper and
 PT textiles production or processing
 PS
 PS Claim 1; Page 45-47; 51pp; English.
 XX
 CC This sequence is the cellulohydrolase CBH B of the invention.
 CC The CBH proteins are used to degrade cellulose, optionally in combination
 CC with other enzymes. Particularly they are used in production, or
 CC processing, of foods, beverages (by liquefaction of plant cell wall
 CC material), animal feeds (particularly silage or grain-based feeds, to
 CC improve nutrient utilisation, and thus growth rate and/or feed
 CC conversion), wood pulp, paper and textiles. The cellulose in the material
 CC being treated may be degraded before and/or after ingestion. A particular
 CC use is in baking where addition of CBH to flour results in increased loaf
 CC volume and better texture (break and shred/crumbs quality). The new CBH
 CC can be produced free of endoglucanases (which can affect fibre strength)
 CC so are better than crude cellulase compositions for fibre modification
 CC where strength retention is important (textile and paper manufacture and
 CC finishing). Cloned CBH and endoglucanases (from WO9713862) should allow

CC preparation of compositions tailored for particular applications. Unlike
CC known CBH, the new enzymes have a broad pH optimum and are effective at
CC the low pH values used in industrial processes, also they have better
CC thermal stability.

XX Sequence 536 AA;

Query Match 5.5%; Score 237.5; DB 20; Length 536;
Best Local Similarity 22.8%; Pred. No. 1.7e-07;
Matches 121; Conservative 66; Mismatches 186; Indels 155; Gaps 23;

QY 348 YSISTPKAPWIKNNFIDVTSEBPSDGLIRLGMISLEIDPTDSNHLMTYGTMTFGCH 407
DB 103 YGVTTSGSE-LRLNFV-----TQSSKNIG---SRLYMSDSSNYELF----- 141
QY 408 DLNMDTRHNVSIOSLADIEEFSVODLASAPGSELLAAGDNGFTFASRNDLGTSPQ 467
DB 142 KLLGDEFTDVDSNLPCL-----NGALFYAMADAGTSEYSGNKAQAK-- 187
QY 468 TWMAPTMTATSTSVYAGNSVKSVMVAVGNTAG-----TOVAISSDGATWSIDYADPT 520
DB 188 -----YGTGYCDSCPRDLKFINGEANCODGWEPSNNVNTGVGDHSGCAEMDVMEAN 240
QY 521 SANGCTVAV-----SADGPTIIMSTASSGVORSQFQGSFASVSLPAGAVIASDKKT 572
DB 241 SISMAFTAHPCDSVSQTMCDGDCGCTGYSASG---DRYSGTCD-----PDGCDYNFRLG 292
QY 573 NSVFYAGSGSTFYVSKDNGSSFFRGPKL-----GSAGTIRDIAHPTTACTLYVSDVGI 627
DB 293 NIDFT-GPELIV---DINSPIFTVOTFITDDGTSSGTLLEIK-----RLVQNGEYI 340
QY 628 FR--STDSCFTFGOVSTALNTFYQIALG-----VGSQ-----SNMNLV 663
DB 341 ANGASITSSVNSGSSITSAFCESSEKTLFGDENVFDKRGLEGMEAMAKGVLVLSLMDXY 400
QY 664 A-----FGTGPSCARLYASDGSASWT--DIQSGSGFSSIDSTKYVAGSGSTA 708
DB 401 AADMMLDSDYPNSSASPTGVAKGTCTDSCGPATVEAESPNAYVTYSNINKEPIGST- 459
QY 709 GGVYVGTNGRGVYAGTGVGGTGGTSSSTKOSSSSTSTLRSVSVSTRASTYT 768
DB 460 -----YSSGSSSG-----SGSSSSSSTTKAISTTLK-----T 488
QY 769 SSRTSSAAGPTSGVAGHYAAGCGIGMTGPTQCAVYVOCOKNDYVYQCV 818
DB 489 TSTTSSGSSSTSAQA--YGCQGGQGMGTPTTCVSGYTCTYENAYYSQCL 536

RESULT 6
AAB47783
ID AAB47783 standard; Protein; 526 AA.

XX AAB47783;

DT 13-MAR-2002 (first entry)

XX Chrysosporium CBH1.

DE Glycosyl hydrolase; family 7; family 10; CBH1; xyl1; fermentation;
KM promoter; terminator; glyceralddehyde phosphate dehydrogenase; GpD1.

XX Chrysosporium sp.

XX Key Location/Qualifiers

FT Peptide 1..19 /label= signal_peptide

FT Protein 20..526 /label= mature_protein

FT Misc-difference 249 /note= "Encoded by ACC"

FT Misc-difference 365 /note= "Encoded by TTN"

FT Binding-site 496..526

/label= Cellulose_binding_domain

XX MO200179507-A2.

XX 25-OCT-2001.

XX 17-APR-2001; 2001WO-NL00301.

XX 13-APR-2000; 2000EP-0201343.

XX (EMBL/) EMBL:AFB M A.

XX Email:farb MA, Punt PJ, Van Zeeijl CMJ;

XX MPI; 2002-066369/09.

XX N-PSDB; AAI72045;

PT New glycosyl hydrolase family 7, glycosyl hydrolase family 10 and
PT glyceralddehyde phosphate dehydrogenase genes from the filamentous
PT fungus Chrysosporium useful for the microbial production of these
PT proteins -

PS Claim 1; Page 34; 43pp; English.

CC This sequence shows a Chrysosporium glycosyl hydrolase family 7
CC protein, CBH1. The CBH1 nucleic acid is used for the industrial
CC production of CBH1 protein by microbial fermentation. The CBH1
CC regulatory sequences (promoter and terminator) are useful for
CC expressing heterologous polypeptides in microbes.

XX Sequence 526 AA;

Query Match 5.3%; Score 228.5; DB 23; Length 526;
Best Local Similarity 24.3%; Pred. No. 6.7e-07;
Matches 118; Conservative 48; Mismatches 150; Indels 169; Gaps 24;

QY 413 DTRHNYSIOSLADIEEFS-----VODLASAPGSELLAAGDNGFTFASRNDLGTSPQ 468
DB 132 DTRKYM-FQLGN---EFTFDVDSNLCGLNGLALFVSMADAGSKYSGNKAQKYG 187

QY 469 VMA-----TPTWATSTSVYDAGN-----SYKSVRVGNATGTOVALS 505
DB 188 GYCDSCPRDLKFINGEANVENMOSSTNDANAGTGRYSGCCSEMDYMEANN-----MAAA 242

QY 506 SDGAGATWSIDYA-ADISMNGGTVAASADGPTILMSTASSGVORSQFQGSFASVSLPAGA 564
DB 243 FTHPCWVIGQSRCEGDCGCT--YSTD-----RYAGICD-PDGC 279

QY 565 VIASDKRTNSVFYAGSGSTFYVSKDYGSSFTRGPKL--GSAGTIRDIAHPTTAGTLVYS 622
DB 280 DFNSTRQGNKTFY-GKGMIV---DTTKKITVVTQFLKNSAGLSLEIKRFRVQNGKAVIPW 334

QY 623 TDVGIRFRTDSC-----TTFGOVSTALNTFYQIALGVGSG-----SNW 660
DB 335 SE-----STIPGEGNSITQDWCROKAAFGDVTDXQDKGMQVOMKALAGPYVLWSTW 389

QY 661 NLAYAF-----GTGPGSARLYA-----SG-----DSGASMTDIQSGSGFG 694
DB 390 DOHAVMNMLDSTWPLDGAACKPGAERAGACPTTSGVAENVAEAPNSVFNIR-----FG 445

QY 695 SIDSTFVAGSGTAGOVYVGTNGRGVYAGTGVGGTGGTSSSTKOSSSSTSSASSSTTL 754
DB 446 PIGST-----VSLGPLDGSAGNPNPV---SSSTPVPSSTT- 478

QY 755 RSSVSVSTRASTYTSSRTSSAAGPT--GSGVAGHYAAGCGIGMTGPTQCAVYVOCOKNDY 813
DB 479 -----SSGSSGPTGTGVAKHYECCGGIGTGPQCESPYCTRLNDM 521

QY 814 YVQCV 818

DB 522 YSOCL 526

```

RESULT 7
AAB82124
ID AAB82124 standard; Protein; 368 AA.
XX
AC AAB82124;
XX
DT 12-JUL-2001 (first entry)
XX
DE Acetylxylian esterase.
XX
KM Acetylxylian esterase; acid resistance; high temperature resistance;
XX animal feed.
XX
OS Chaetomium gracile.
XX
FH Key Location/Qualifiers
FT Peptide 1..25
FT /label= Signal_peptide
FT Protein 26..368
FT /label= Mature_protein

JP2001054383-A.
PD 27-FEB-2001.
PF 16-AUG-1999; 99JP-0229769.
PR 16-AUG-1999; 99JP-0229769.
PA (NIPC) SHIN NIPPON KAGAKU KOGYO KK.
DR WPI: 2001-268946/28.
DR N-PSDB: AAF86576.
XX
PT New protein with acetylxylian esterase activity comprises acid and high
XX temperature resistance.
XX
PS Disclosure: Page 9-11; 14pp; Japanese.
XX
CC The present sequence is acetylxylian esterase, which has acid and high
XX temperature resistance. The protein of the present invention is derived
XX from Chaetomium gracile. Acetylxylian esterase can be used for animal feed
XX composition together with the other xylan decomposing enzymes.
XX
SQ Sequence 368 AA:
Query Match 5.2%; Score 227.5; DB 22; Length 368;
Best Local Similarity 23.8%; Pred. No. 4,9e-07;
Matches 98; Conservative 54; Mismatches 162; Indels 97; Gaps 16;

OY 432 VODLASAPGSGELLAAGDNGFTFASRNDLGTSPQTVNATPTMATSTSYDVAGNSVKSIV 491
DB 31 VTFGSGNPGSARMYIVP-----KLGSPAILFVAVHCTGTANAFYGT---PY 77
OY 492 VRGNTAGTQVAL--SSDGGATSIDYADTSNNGCTVAVSADGDTILNASTASGVORSO 549
DB 78 ARLADYGFIVYEPESPHDGGCWDVSSRAITLTHNG----- 113
OY 550 FQGSFASVSLPAGAVIASDKTNSVFYAGSGSTFYVSKDPTSGSFTRGPKLGSAGTIRDI 609
DB 114 --GDSNSIANMYTTINQYKADRNRFVACTSSGAMTNVLSITY--PDLFPAAS----- 164
OY 610 AAHPTTAGTLVYSTDVGFIFRSTDSGTFEGOVSTALNTVOIALGVSGSMNLVAFGTGP 669
DB 165 AVAGVPAGCGFYTAVG---WMDTCAGQGVYTSQAMNQFAL-----NMYPGYTCP 212
OY 670 SGARLYASGDSGAS-----WTDIQSGQFGSISTVVASGSA-----GOV 711
DB 213 RPAMLLIHHSADTTIYPONFNETLKOMAGVIG-YTYGPOPOOT-LPGNPSESEYKYVGP 770
OY 712 YVGTNGRGVYAGCTVGCGF---GGTSSSTKSSSTSSASSSTLRSSVSTTRASTV 767
DB 271 LVGIYGTGITHNIPVDGAGMEWFGITGSST-PTTSTAGCTTSTKSS----- 319

```

```

OY 768 TSSRTSSAGPTGSGVAGHYAOCGGIGMTGPTQCVAPVYCOXONDYVYOCV 818
DB 320 -STSTAPTSGPSGC-TSARWQCGGIGMTGCTVCAASPMKCTYANBMWYSCCL 368

RESULT 8
AAR80279
ID AAR80279 standard; Protein; 516 AA.
XX
AC AAR80279;
XX
DT 21-DEC-1995 (first entry)
XX
DE N. crassa cellobiohydrolase-1.
XX
KM Cellobiohydrolase-1; CBH-1; cellulase complex; promoter;
XX cellulose binding domain.
XX
OS Neurospora crassa.
XX
FH Key Location/Qualifiers
FT Peptide 1..28
FT /note= "Sig_peptide"
FT Domain 29..470
FT /note= "catalytic domain"
FT Region 471..519
FT /note= "hinge region"
FT Domain 520..550
FT /note= "C-terminal cellulose-binding domain"
FT Misc-difference 1..516
FT /note= "these features, given in the
FT disclosure, do not correspond to the
FT number of amino acids encoded by the
FT CBH-1 gene sequence, given in Fig.1"

W09519441-A1.
PD 20-JUL-1995.
XX
PF 11-JAN-1995; 95WO-GB00049.
PR 14-JAN-1994; 94GB-0000623.
XX
PA (UYLE-) UNTV LEEDS.
XX
PI Parish JH, Radford A;
XX
DR WPI: 1995-263871/34.
DR N-PSDB: AA098564.
XX
PT New regulated promoter from Neurospora crassa cellobiohydrolase-1
XX gene - and constructs etc. contg. it, and transformed filamentous
XX fungi, provides increased prodn. of cellulolytic enzymes or of
XX heterologous peptide(s), e.g. hormones
XX
PS Disclosure: Fig.1; 32pp; English.
XX
CC The N. crassa cbh-1 gene was isolated by probing a genomic library
XX with a PCR product generated using primers based on known CBH
XX sequences. The regulated promoter of the gene can be used to
XX achieve high yields of glycosylated heterologous proteins in N.
XX crassa, and the gene can also be used for CBH-1 manufacture.
XX
SQ Sequence 516 AA:
Query Match 5.2%; Score 226; DB 16; Length 516;
Best Local Similarity 26.1%; Pred. No. 9,6e-07;
Matches 78; Conservative 37; Mismatches 88; Indels 96; Gaps 11;

OY 563 GAYIASDKTNSVFYAGSGSTFYVSKDPTSGSFTRGPKL--GSGTIRDIAAHTTAGTLY 620
DB 271 GCDPFSYRMGNITFY-GECKTV---DTSSKFTVVTQFIKDSAGDLAEIKAFYVONGKYI 325

```

Oy		621	VSDVGIERNDSG-----TTFG-----QSTALNTYQIALGSGSS	658
Dd		326	ENSSNNDVGVSNGNSTTOSFCKSQKTAFAGDIDDENKKGLKOMGAQAQAVLWST----	381
Oy		659	NWNLYAF-----GTGPGARLYASGDGSAMWTDIQSGCGFSIDST	699
Dd		382	-WDHRAANMLMLDSTPYYPVKVPGAYRBSGPPTTSVPADVDANA-----PNS	426
Oy		700	KVAGSGTAGGVYVTGNRGVFYAOGTVGGTGCTGSTSSSTYOSSSTISSASSSTLLRSSVV	759
Dd		427	KVAFSNIKFIHGHLGI-----SPFGSGSSTPPSPNPSSSASPSS-----	463
Oy		760	STTRASTYTSSRTSSAAGPTGSGVAGHYAACGGIGMTGPGQCVAAPYVCOKONDYYOCV	818
Dd		464	----STAKPSTSTASNPSTG- <u>AAHMAOCCGIGFSGPTTCPEPYTCARDHDYISQCV</u>	516
RESULT		9		
ABBO5058				
ID	ABBO5058	standard; Protein: 526 AA.		
XX				
AC	ABBO5058:			
XX				
DT	27-MAR-2002	(first entry)		
DE				
XX	Trichoderma reesei cellobiohydrolase I (CBHI) 55KD (family 7) protein.			
XX				
KM	Trichoderma reesei; filamentous fungi; phenotype; characterisation;			
XX	fermentation; screening; morphology; cellobiohydrolase I; CBHI.			
OS	Trichoderma reesei.			
XX				
FH	Key	Location/Qualifiers		
FT	Peptide	1..19		
FT		/label= signal		
FT	Protein	20..526		
FT		/label= cellobiohydrolase_I		
FT	Misc-difference	249		
FT		/label= unknown		
FT		/note= "encoded by NCC"		
FT	Misc-difference	365		
FT		/label= unknown		
FT		/note= "encoded by TTN"		
XX				
PM	WO200125468-A1.			
XX				
PD	12-APR-2001.			
XX				
PF	13-APR-2000; 2000WO-USI0199.			
XX				
PR	06-OCT-1999; 99WO-NL00618.			
XX				
PA	(EMAL/) EMALFARB M A.			
XX				
PI	Emailfarb MA;			
DR	WPt. 2001-281733/29.			
DR	N-PStDB; ABA927722.			
XX				
PT	Expressing heterologous proteins encoded by a library of DNA vectors,			
PT	involves stably transforming mutant filamentous fungus with the vectors			
PT	and culturing transformed fungi for expressing heterologous proteins -			
PS	Disclosure: Page 66-69; 85pp; English.			
XX				
CC	The present invention describes a method of expressing a number of			
CC	proteins encoded by a library of DNA vectors (I). The method involves			
CC	stably transforming a mutant filamentous fungus (II) with (I) so as to			
CC	introduce into each of a number of individual fungi, at least one			
CC	heterologous protein-encoding nucleic acid sequence (III), and culturing			
CC	the transformed mutant filamentous fungi for the expression of			
CC	heterologous proteins encoded by (III). (i) comprises a number of			

Query Match	5.2%	Score 224.5	DB 22	Length 526
Best Local Similarity	22.0%	Pred. No. 1.2e-06		
Matches 121:	Conservative 64	Mismatches 169	Indels 195	Gaps 27
XX	Sequence	526 AA:		
CC	different vectors, each comprising a different protein-encoding nucleic			
CC	acid sequence being operably linked to an expression regulating region			
CC	and optionally a secretion signal encoding sequence. (ii) has a phenotype			
CC	characterised by a culture viscosity, when cultured in suspension, of			
CC	less than 200 cP at the end of fermentation when grown with adequate			
CC	nutrients under optimal or near-optimal conditions. The method is useful			
CC	for expressing large quantities of heterologous proteins that are useful			
CC	for isolation, characterisation and application testing, and also for			
CC	commercial production of proteins. The mutant filamentous fungi obtained			
CC	by the method are suitable for high-throughput screening techniques			
CC	owing to their unique morphology and very low viscosity of their			
CC	cultures. The present sequence represents the Trichoderma reesei			
CC	cellobiohydrolase I (Cht1) 55kD (family 7) protein, which is given in			
CC	the exemplification of the present invention.			
XX				
YY	391	DSN -HWLGTGTTTCGGDLTWDTRH-----NVSIQSLADGIEEFVSODLASACGSGSL	444	
DB	52	DANMWTHTRTD-SATNCEYEGKNKMDTSYCSDPGSCASKCCIDADVSTYGIITS-----	104	
YY	445	IAAVDDNCGTFPASHRNDLCTSPQTWAFPTWATSTASVDY-----AGNSVKSVARGNT--	497	
DB	105	-----GNSLNLKRYTKGQYSTN-----IGSTFTYIMESDFTKYMQLLGNETTFPVDSNLGC	156	
YY	498	---AGTOVAISSDG-----GATWSIDY-----AADTS	521	
DB	157	GLNGALYFESPMADGCMGSKYSNGKAGAKYGCYCSOCPRDLKFTNGEANNVENMOSFTWD	216	
YY	522	NMGGAIVAY-SADGDITLLMTASSG-----VORSQFOGS-----FASVSSL	560	
DB	217	ANAGTGKIGSCCSEMDWEANMAAFTPHPCXVIGQSCSECDSCGGITSTRYAGICD--	275	
YY	561	PAGAVIASDKRTNSVFYAGSGSTFYVSKDTGSSFTFRPKL--GSAGTTIRDIANHPTTACT	618	
DB	276	PGDCPFNSYRQGNKTFY--GKGMTV---DTTKKITYVQFLKNSAGELSEIKRFYQNGK	330	
YY	619	LXVSTDVGFRRBDSG-----TTFCQVSTALNTVQIALGVSG-----	657	
DB	331	VLPNSD-----STPGEVBSNSTIDQMCBQRKAFFGDYVDXQDKGGMVQKALACPMVLV	385	
YY	658	-SNMNLVAF-----GTGPGSARLYA-----SG-----DSGASWDDIQGS	690	
DB	386	MSIMWDHAANMLMDLSTWPIIDCAGKPAERGCAPTTSGVPAVEAEAPRASNVIFFSNIR--	443	
YY	691	QGFSGSIDSTKKVAGSGSTAGQYVVTINGRCVFAAGTGVGGGTGCTSSSTKQSSSSSTSSASS	750	
DB	444	--FGPIGST-----VSGIIPDDSGSNPNPNPV---SSSTPVPSS	475	
YY	751	STTLRSSVYSTRASVTSSPTSSAAGPT--GSGVAGNHAYQACGIGTGTQCVAPRYVCK	809	
DB	476	STT-----SSGSSGPIGTGTGVAKHNYDQCGGIGTGTQCESPTTCK	517	
YY	810	QNDYYQCV	818	
DB	518	LNDWYSQCL	526	
XX	RESULT 10			
XX	ABB60186			
XX	ID	ABB60186 standard; Protein; 1714 AA.		
XX	AC	ABB60186;		
XX	DT	26-MAR-2002 (first entry)		
XX	DE	Drosophila melanogaster polypeptide SEQ ID NO 7350.		
XX	XX	Drosophila; developmental biology;cell signalling; insecticide;		

CC One example of a cellulose-degrading enzyme of the invention.
CC See AA014838-Q14841.
XX
SO Sequence 617 AA:
Query Match 5.18; Score 219.5; DB 12; Length 617;
Best Local Similarity 21.9%; Pred. No. 3.3e-06;
Matches 149; Conservative 81; Mismatches 247; Indels 203; Gaps 32;
OY 242 KAKIQAPEAKLYLTSGTGPYDGTGLGVMRYDAGTGMKDIFFVSGSD-LYF-----G 294
DB 38 RAAISP---NITGNOLSGTENNSSRLGKRLTGYNWENMSSACRDLNHSDDLCC 94
OY 295 FGGL-GLDLQKPGTLVVASLNSMMPAQLFRSTDSGTTWSPIMAMASYPTETYSIST- 352
DB 95 NGVDPDTCCKPGAVVYAFHD-----KSLQNGA-----YSIVTL 128
OY 353 PKAPMT---KNFTDIVSESPDGLIRLGMIF-----SLEIDPTDSNHL----- 396
DB 129 QMAGYVRDKNPGVDESETPAPSPMDK-----VEFAKNAPFSLDPDLNDQOVYDEEVNF 183
OY 397 ---YGTGMTIFGGHDLTNMDTRHNVSIOGLADIEEFVSODLASAPGSELLAAVGDND 452
DB 184 LVNRKYNAST-----STGIRKAYSL-----DN 204
OY 453 GTFPASNDLGTSPQVWATPTMATSTSVYAGNSVSVVAGNTA---GTQVAT--- 504
DB 205 EPALMSTHRIHPEOLOAEELVAKSIDLSKALKNDPHEIIFGPALYGCAYLSLDAP 264
OY 505 ---SSDGLATMSIDYADTSMNGSTVAYASADGDTIL-----WSTASSGVQRSGFQGS- 553
DB 265 DWPSLQGNYSWFIDYLDQMKN---AHTONGKRLDLVDVHWPEAOGGGQRIYFEGAG 320
OY 554 -----FASVSLPAGAVIASDKTNSVFVAGS----- 580
DB 321 NIDTKARVQAPRSLMPAYOEDSMIGTWFSY--LPLIRKLOSSIQT---YVGTGLAI 375
OY 581 -----GSTFVSKDPTGSSFTRGPKLSAGTIRIDIAH--PTTAGTLYVSDVGIESTDS 633
DB 376 TERFSGGDNHIS---GGLTADALGIFGKYGVYAANYMOTEDNTDYSAAVKLYRNVDG 431
OY 634 GTT-FCGV-----STALNTQIALGVGSGSNMNLXAFGTG---PGCALIXASGD----- 679
DB 432 NKGGFGSIKDAATSDTENSSTVASTDEENSELHLIYLNKNEPDPINATFOULSGDKTYT 491
OY 680 SGASMTDIOGSGFGSIDSTRVAGSGSTAGOVYVGTNGRGVFAOQGVG-CTGCTSSST 738
DB 492 SGRW-----GF---DQT---GSDITEQAALITNINNNQFTYTLPLLSAVHILKADST 538
OY 739 KOSSSTSSASSSTTLASSVSTTRASTVSSRTSSAAGPTGSGVAGHYAOCGGIGMTGP 798
DB 539 EPISEIPISSSTSPVQPTSTSTSTSTSTSPVQPTTSPSCG-TABRMAOCCGNGMSCC 597
OY 799 TOCVAPYVCOKONDYVQCV 818
DB 598 TTCVAGSTCTKINDWYHQL 617
RESULT 13
ID ABB54925
ABB54925 standard; Protein; 1648 AA.
XX ABB54925;
XX
XX 16-MAY-2002 (first entry)
XX DE Lactococcus lactis protein yqbk.
XX
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX Lactococcus lactis IL1403.
XX
XX
XX FR807446-AL.
PN

XX
PD 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Bolocrine A, Sorokine A, Renault P, Ehrlich SD.
XX
XX WPI; 2002-043418/06.
XX
XX
XX New nucleotide sequence useful in the identification of Lactococcus
XX lactis and related species -
XX
XX Claim 6: SEQ ID No 1627; 2504bp; French.
XX
XX The present invention is related to a Lactococcus lactis nucleotide
XX sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
XX nucleic acid sequence is useful in the detection and/or amplification of
XX nucleic acid sequence, particularly to identify Lactococcus lactis or
XX related species. The proteins of the invention are useful for the
XX biosynthesis or biodegradation of a composition of interest. The
XX invention helps research in lactic bacteria, particularly useful in the
XX production of yogurt and cheese.
XX Note: The sequence data for this patent is based on equivalent patent
XX WO200177334 (published 18-OCT-2001) which is available in electronic
XX format directly from WIPO at http://wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1648 AA:
Query Match 5.0%; Score 218; DB 23; Length 1648;
Best Local Similarity 22.0%; Pred. No. 1.5e-05;
Matches 220; Conservative 120; Mismatches 324; Indels 338; Gaps 60;
OY 29 AYATDITG---GLYRLNADSWTAVTGDIAD-----NAGMHNKIDAVALDPDDQKY 78
DB 124 SYPTDAGVOOALYAM-----YSATNASTSDPGLYYGANTLNSLTKAVSNPSATNMT 178
OY 79 YAAVGMVYTN--WDPSNGLIRSSDRGATWSPNLPKVGKNNPGRAGERLAVDPANSN 136
DB 179 FSTLKHAKSLTW-ISPAPDLTSSNSOPSGTNYT---GTLP-----SN 219
OY 137 IYFG-----ARSG-----NGLW-----KSTDGVTEKSVSPFATGT 169
DB 220 -YFEGVPTIRNVTVAASGDNITVAQNAFATNGSMITGAPNITGCGINDSDIG--NTNL 276
OY 170 YIPDPSDSNGY-----SDKGLMWTFFDSTST---TGANTS----- 204
DB 277 YIGATGSIAGMINIYGCNAAATISGNTHTVIAOSSSTINSVTGSSAGTITISGNTMLDIS 336
OY 205 -----RIF---VGTAD---NITASV--YV--STNAGSTWASVAPQPGKYPFHKKLQP 247
DB 337 GATASQITNITYGAGIGTNSPVAVNGVNTTYVNSTNG----- 374
OY 248 AEKALYLYSDGTGPDGLGSMRYDIAGTW-----KDTTPVSGSDLYFGFGL--GLD 301
DB 375 ---ARYOLYOGGT-VYGNISGSIYNTLSGAGMTGATSNINGAGPASPNGSGFCGNIG 430
OY 302 LQKPGTLVVASLN--SWMPAQLFRSTDSGTTWSPIMAMASYPTET-----YYSI 330
DB 431 TSGAGNIVSNYNTSSFTGQALFTGNACT-----GASYAOATNSTTAAGILLIYANT 483
OY 351 STPKAPWIKNFIQVTSSESPDGLIRLGMIESLEIDPDSNHLMLYGT--TGM-TIFGCH 407
DB 484 TN---YIKSAFTTGTAGAVY-GIVGNGH--DSLKISP---SOMGLSGGTGLDSAVGT 533
OY 408 DLTNMDTRHNVSIOGLADIEEFSVOD-----LASAPGSELLAAVGDNDGFTFA 457
DB 534 DAKAYGQIPSTVVSNAOKITSNALIGDYVTMLQSGVMSTTSG-----GANDMTGYSVG 588
OY 458 SRNDLGTSPQVWATPTMATSTSVYAGNSVSVVAGNTAGTQVAISSDGLATMSIDYA 517


```

Db 589 ASSNGYLOGSVLEAGTANADNSVGAG--VYCSAMGTAETTLA-----AOGTVAYG 640
QY 518 -ADTSM-----NGTV-----AYSADGDTIL-----WSTA--SSGVQRQFQ 551
Db 641 KPTSSAINSQMDLWGGGTWMTYRQAFQFQNGNSYLLHNNDIARMTYGGQNG--S0VG 697
QY 552 GSF-----ASVSSLPAQAVIASDKTNSVFYAGSGS--TFYVS-----KDTGSSFTFR-- 596
Db 698 NSVNLNGALVDLLEGGYATITKMGNTTAOVNOGOVNMFLSGSGMDLNTGSAITNVY 757
QY 597 -----GPKLSAGTIRIDIAHPPTAGTLVYSTDVGIERSD-----SGTFEGOV 640
Db 758 NGYINATTCGNYOAGV-----ETIAGDSTVNVYGDGSGSPRTGKQLCGPFENG 810
QY 641 STALTN-----YQIALG-----VSGSWMNL-----YAF 665
Db 811 SSLGNTALNVDLTGSGSFQLPSPGYISGAGYNTVTVHVGSVNNSISVISANAAS 870
QY 666 GTPSGARLYASGD--SGASWTDIOGSGFSGIDSTKRVAGSGSTAGQVYVGTN----- 716
Db 871 GNVLNGAVITGDDGSGTSGNSTYTNV-----GTINMT-INADGTVGSYV-ATVYVAMPA 922
QY 717 -----GCVVYAGTVGGGT--GCTSS-----STKQSSSTSSASSST-----TLRSS 757
Db 923 SGGRYNTNIGIGDTTISGTTSGSSDNLTDIAVANNKNSAITLGNSTSHNPITINGS 982
QY 758 VVS-----TTTASTVSSRTSSAGPTGSGVAGHYAOCCGI 793
Db 983 LNNNSAETIEKAVNVVAGSFKMGGATANNHAATYSKHGSI 1024

```

RESULT 14

AAB23367
ID AAB23367 standard; protein; 1160 AA.

AC AAB23367;

DT 09-JAN-2001 (first entry)

DE Mouse trophinin protein.

KW Trophinin; vaccine; contraceptive; mouse; human; chromosome X.

XX Mus sp.

OS key Location/Qualifiers

FT Domain 1..50 /note="Cytoplasmic domain at the amino terminus"

FT Domain 187..335 /note="Cell surface domain"

XX US611089-A.

XX 29-AUG-2000.

XX 28-FEB-1997; 97US-0808599.

XX 04-OCT-1994; 94US-0317522.

XX 12-MAY-1995; 95US-0439818.

XX (BURN-) BURNHAM INST.

XX Fukuda MN;

XX WPI; 2000-586380/55.

XX N-PSDB; AAA91976.

XX Novel nucleic acid sequence encoding mouse trophinin, for producing

XX trophinin vaccine useful to prevent pregnancy in an individual by

XX eliciting an immune response which blocks embryo implantation -

XX Claim 1; Fig 1; 57pp; English.

XX
CC This present sequence is the mouse trophinin protein. The DNA encoding
CC this sequence was isolated from a genomic mouse library using human
CC trophinin cDNA as a probe. Mouse trophinin is used as a vaccine for
CC preventing pregnancy in an individual by eliciting an immune response
CC against trophinin that blocks embryo implantation. Due to the limited
CC expression of trophinin, an anti-trophinin immune response is less
CC toxic than other prior art anti-pregnancy vaccines directed to peptide
CC hormones.

SQ Sequence 1160 AA:

Query Match 5.0%; Score 217.5; DB 21; Length 1160;

Best local similarity 22.0%; Pred. No. 1e-05;

Matches 187; Conservative 84; Mismatches 357; Indels 223; Gaps 36;

```

QY 6 NVKLGGGGFVPGIIEHPKTKGVAVARTDIGLYRLNADSWTAIVYDGIADNGWHNMG 65
Db 469 SVSFGGGSSGANAGF-----CGTLNSTSPFG-----AISTGFGSALNNSA---NFG- 513
QY 66 DAVALDPQDDQKYAAVGYMTNSMDPSNGALIRSSDKGATW---SF-----TNLPKVG 116
Db 514 -----GAISTSGVLSNSASRFGALINTSAGFSLNSASRFGALSTSASF--G 561
QY 117 GMPGR-GAGERLAVDPANSNIIFGARSGN-GLMKSTDGVTF-SKVSFTATGYIPD 173
Db 562 GVLNGRAGFCGAL-----NTNATFGVLSNSAGFCGAMNTNATFGALNSNAGFCALST 616
QY 174 PSDSNGYNSDKGLMWTTFDSTSTTGA--TSRIEFTGTDN--ITASYVTNAGSTW 228
Db 617 STNFGGALNNSAGF-----GGAMNTSASFEGVLSNSAGFCALNTSASFEGA- 663
QY 229 SAVPGQPKYFPKAKLQPAEKALYLTFSDGTGTYDGLSVNRHYDAGTMDITPVSG 288
Db 664 -----LTNSAGFCALSTSASF-----GALNNSA----- 688
QY 289 SDLYFEGGLGLDLQKPGTLVVASLNSWMPDAQLFRSTDGTTWSPYMAASYPTERY 348
Db 689 -----GFGG-----AISTSASFEGALNNSAGFCG 712
QY 349 SISTRKAPWIKNNFIDVTSPPSDG---LIRLGNHIELEIDPTDSNMWLTGNTT- 403
Db 713 AIST-----NASFGAISNSPFGAFSTSVGFGGLTNTDGSNHSNISISFSAFTTS 766
QY 404 --FGCHDLTNMDFRHNSIQSLADGIEFVSVDLASAPGSELLAAGDNGFTFASRND 461
Db 767 VSFGGSHSTN-----LCFG-----GAPSTSLCEGASNTN-LCFGGSN- 803
QY 462 LGTSPQTVWATPTWATSTVDY-----AGNSYKSVYRGNTAGTOVAISSDGGATWS 514
Db 804 -----STNCFSGATSANFNNGHSISFGNGISTSAGCGNGLGTSAGFDSLSLTGTF 854
QY 515 --DYADTSMNGTVAVSADGDTILNSTASSG--VORSQFQGSFASVSLPACAVIASDK 570
Db 855 GGSIGPSASFNGLGTSTGCGGGLGTSTDFSGGLNHNADNGLGNSAGFNGGL----- 908
QY 571 KTNVSVEYAGSGSTFYVSKDQSGSFTRGPKL-GSAGTIRIDIAHPTTNGTLVSTDVGI 629
Db 909 NTNIDFGEGLGTSAGFCDGAGSSTIFGAGLVTSDFGAFAGNLGTNGFGCTL--GTGAGFSV 966
QY 630 STDGTFEGOVSTALNT-YQIALGVSGSNMNL-----YAFGTGSGARLYASGD 679
Db 967 SLNNGNGFNGPNASFRGLNTIIGFSGSNTSNGFTGEPNTGSSFFNGSSSLVGFSGCP 1026
QY 680 S-GASWTDIOGSGFSGIDSTKRVAGSGTLAGVYVGTNGCVFYAGCT-----VGGG 730
Db 1027 STGAGFCGSPSTGFGGPGFPSTGPGFPGFPSTGPGFPGFPSTGPGFPGFPSTGPG 1086
QY 731 TCGRSSSTKQSSSTSSASSSTLRSVSTTASTYTSRTSSAAG-----PTGSGVA 784
Db 1087 FCGPSTGCGFGGPGSTGCGFPGSTAGFSGSLSTFGGGLNTSAGFSGSPSTGTGFG 1146
QY 785 GHYAOCGGIGW 795

```


FILING DATE: 12-MAY-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 24:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1160 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-808-599A-24

Query Match 5.0%; Score 217.5; DB 3; Length 1160;
 Best Local Similarity 22.0%; Pred. No. 5.7e-07;
 Matches 187; Conservative 84; Mismatches 357; Indels 223; Gaps 36;

6 NVKLGCGGCVPIIHPKTKGAVARTDGLYRLNADDSMTAVTDGADNAGMHWI 65
 469 SVSFGSSCANAGF-----GCLNASTSFG-----AISTSGSALNSA---NFG- 513
 66 DAVALDPODDQKYAAVGMTNSMDPSNGAIIRSDRGATW---SF-----TNLPKYG 116
 514 -----GAISTSGSGLNNSASFGALINTSAGFSTLNSASFGALSTASG--G 561
 117 GNNPGR-GAGERLAVPANSNIITFGARSGN-GLMKSTDCGVTF-SKVSFTATGYIPD 173
 562 GVLNGRAGFGAL-----NTNATFGVLNNSAGFGGAMNTNATFGALNSNAGFGAIST 616
 174 PSDNNGYNSDKOGLMWTFFDSTSTGGA--TSRIFVGTADN---ITASYVSTNAGSTM 228
 617 STNFGALNNSAGF-----GAMNTSASFGCVLNSAGCGALINTANFGA- 603
 229 SAVPGPGKTFPHKAKLQPAEKALYLYSDGTGPDGTLGSVMRYDIAGGTWKDITPVSG 288
 664 -----LTNSAGFGAISTASG-----GALNNSA---- 688
 289 SDLYFGFGGLGDLQKPGTLVNASLNSMWPDALFRSTDSGTWSPIMAASTPTEYYY 348
 689 -----GFGG-----AISTSGSALNNSAGFGG 712
 349 SLSPTKAPWIKNNFIDVTSSEPSDG-----LIKRLGMWIESLEIDPDSNHLWTGWTI- 403
 713 AIST-----NASFGALNSPDPGCAFSTSVFEGGTINTDGCNSHNSISFGSNPTTS 766
 404 --FGHDLTNMDTRHNVSIQSLADGIEERSVODLASAPGGSSELLAAVGDNDGFTFASRND 461
 767 VSFSGSHSTN-----LCFG-----GAPSTSLCFGSSASNTN-LCFGGSN- 803
 462 LGTSPTVATPTWATSTVDY-----AGNSVKSIVRVKNGAGQVVALSSDGCATWST 514
 804 -----STNCFSGATSNFNEGHSISFNGNLSTAGFNGLGTSGAGDSSSLGTSTGF 854
 515 --DYAADTSMNGCTVAYSADGDTILMSTASSG--VORSQFGSFAVSSLPAGAVIASDK 570
 855 GCSLGSASRNGSLGTSTGCGGLGTSTDESGGLNHNADNGLNLSAGNGL----- 908
 571 KINSVFYAGSGSTFYVSKDTGSSFTFRPKL-GSAGTIRIDIAAHPPTAGTLVYSTVDGIFR 629
 909 NTMTDGEGLGTAGFGDGLSGSTSFAGALVYTSDFAGNLGTNTGFCGTL--GTGAGFSV 966
 630 STDSGTFPGOVSTALNT-YQIALVGSQSNML-----YATGTGSGARLASCD 679
 967 SLNNGNGFGGPNASFRKGLNTLIGFGSGSNTNGFTGEPNTGSSPSNPSIVGSGGP 1026
 680 S-GASMTDLOGSOGSISDSTKYAGSGSTAGQVYVGTNGRGVFAAGT-----VGCG 730
 1027 STNAGCSGSGSTGFGGPGSTGFGGPGSPGPFGGPSTGCGGFGCGPSTGGG 1086
 731 TGGTSSSTKSSSTSSASSSTTLKSSVSTRASTVTSRTSSAAG-----PTGSGVA 784

Db 1087 FGGPSTGGGFGPGSTGGFGGPGSTAGFGSGLSTSTGFGGLNTSAGFGGPPSTGTGFG 1146
 QY 785 GHYACGGCIGW 795
 Db 1147 GGASSHGCGCF 1157

RESULT 7

US-08-194-290-7
 Sequence 7, Application US/08194290
 Patent No. 5500353
 GENERAL INFORMATION:
 APPLICANT: Smith, John
 TITLE OF INVENTION: Bacterial surface protein expression
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Shlesinger, Arkwright & Garvey
 STREET: 3000 South Eads Street
 CITY: Arlington
 STATE: Virginia
 COUNTRY: USA
 ZIP: 22202

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/194,290
 FILING DATE: 09-FEB-1994
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
 NAME: Garvey, George A
 REGISTRATION NUMBER: 17737
 REFERENCE/DOCKET NUMBER: 5946-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-684-5600
 TELEFAX: 703-836-5288

INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1026 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-194-290-7

Query Match 4.7%; Score 202; DB 1; Length 1026;
 Best Local Similarity 21.3%; Pred. No. 5.8e-06;
 Matches 193; Conservative 117; Mismatches 360; Indels 234; Gaps 43;

6 NVKLGCGGCVPIIHPKTKGAVARTDGLYRLNADDSMTAVTDGADNAGMHWI 65
 111 NLTATGAGAG--ATAFPAAYTVSAOT-VATAY-----DKILGNVATAAGV 154
 QY 66 DAVA-----LDPDOD-----DOKVYAA-----VGMVTN 87.
 155 DVAATAAFLSRQANIDYLAFAVRANPTFAAADIDLAVKRAALLGTLINATVSGIGCYAT 214
 Db 88 SW-----DPSNGAIIRSDRGATWS-FTNLPFK-VGNNMPGRGAGERLAVPANSNIIFY 140
 215 ATAMINDLSGAL--STDMAAGVNLFTAVPSSGVSGLSTLTGTGDTLGTANNDTEVA 272
 Db 141 GAASGGLMKSTDCGVTFSKVSFTATGYIIPPSDSNGYNSDKGLMWTFFDSTSTG 200
 273 GEVAG-----AATLTVGDTL-----SGGAGTV--LNMVQAAAVTALPT 309
 QY 201 GAT-----SRIFGTADNTIASVYSTNAGSTSAVPGQPKYFPHKAKLOPA 248
 Db 310 GVTISGLEMTNNTVSGAALTILNTSSGVYGLTALTNTNSGAQOYTTAGACGNLTITTAQAA 369.
 QY 249 EKALYLYSDG-----TGPYDGLGSVMRYDIAGGTWKDITPVSGSDLVFG----FG 296


```
Db 370 NNAV-----DGANVTAVASGVTSGTT-TVGAASASGT-VSVVANSSTTTTGAIVTG 423
QY 297 GLGLDL-OKRGTLVAVASLMSWMPDOLFRSTDSCTTSPIMAMASYETLEYYSISIPKA 355
Db 424 GTATVAVQRTAGNNA-----NTTLOADVTYGNSSST-----AVVTQTPAAGATVA 472
QY 356 PWKNEFDVTSSEPSD-----GLIKRLGMMIESLEIDPTDSNMH-----LYGTGWTI-FGG 406
Db 473 GRV-NGAVTTTDSAAATAGKRIATV--TLGSCGATIDSSALTITVNLGCTGSLICGR 529
QY 407 HDLTNMDTRHNVSIOGLADGIEFVSODLASAPGSELLAAVGDNGFTFASRNDICTSP 466
Db 530 GALTATPTATLT-----NVNGLTTTGAITDSEAAA--DDGFT--TINAGSTA 575
QY 467 QTWAFPTMTATSTSVDAAGSVSVAVGVTAGTQOVAISSDGCATWSIDVAAOTSMNGCT 526
Db 576 SSTIASLVADATTNLISGARVYITSHTAALGTITVTSVGATLGAELATGLVETGG- 634
QY 527 VASADBDTILWSTASSGVORSGFSGFASVSLPAGAVIASDKTNSVFYAGSGSTFYV 586
Db 635 -----AGADSLTLGATTAIWMAGADD-VTVSSATLGA-----GGSVNGGDTIVLV 681
QY 587 SKDTGSFTNGPKLGSAGTIR--DIAAHPPTAGTLVYSTDVGFIRSTDCTTGGQVS--- 641
Db 682 ANVNGSFSADPAFGGFETLRVAGAAQGSNANCGFTALDQ--ATAGATTFTNAVNV 738
QY 642 --TAL-----TNTYQILGVSGSNNL-----VAFGT----- 667
Db 739 GLVLAPPTTITVTLNATGTSVFNLTLSSSAALAGVALAGVTVAIATDTNTTA 798
QY 668 -----GPGSARLYASGDSGASMTDIQSGGFSIDSTKVAGSGSTAGOVYVNGRG 719
Db 799 HVDTLFLQATSAKSIVYTGANGLNTN-TGNATVTSFDASAVGTGSAVFEVSANTTVE 857
QY 720 VFAAGVGGGTGCTSSSTQSSSTSSASSSTTLRASSVSTRASTVTSKRTSSAAGPT 779
Db 858 VV-----TIRGAG-----ADSLTGSATANDTIIGAGADT---LVYTGCDTFTTGT 902
QY 780 GSGV 783
Db 903 GADI 906

RESULT 8
US-08-317-522A-2
; Sequence 2, Application US/08317522A
; Patent No. 5599818
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophobin and Trophobin-Assisting
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,522A
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9991
```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 749 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-317-522A-2
```

Query Match 4.6%; Score 201; DB 1; Length 749;
Best Local Similarity 23.2%; Pred. No. 4.4e-06;
Matches 184; Conservative 86; Mismatches 320; Indels 204; Gaps 36;

```
QY 92 SNGATITS--SDRGATMSFNLNLPKVGANMPGKAGRLAVDPANS-----NIITYFGA 142
Db 42 SRGASTRAGFSDR-ASISFNGAPSSSGF--SGPGITTFVAPSTASFSNTASISFGCT 98
QY 143 RSGNGLMKST-----DGVTFKVSSEFTATGTATIPDPSDNGYNSDKGLMWTFD 193
Db 99 LSTSSSFSSAASISFGCAHSTSTFSSEASISFGCMPTASFSGVSFSFGL----- 133
QY 194 STSSTTGATSRITFVTADNITASVYVSTNAGSTWSAVPGQKYPFHAKKIQPAEKALY 253
Db 154 STSATFSGGASSGFGCT-----LSTTAG--FSGV-----LS 182
QY 254 LITSDBGCPYDGLGWSWRDIAGCTKKDITPVSGSDLYFGGGLGLDLPGLTVVASL 313
Db 183 TSTSPGAPPTTSTVFS-----SALSTSGFGG-----ILST 213
QY 314 NSMWPDLQLEFRSDSGTTPWIPIMAMASYPTET--YYISISTPKAPWKINEFDVTSSEPSD 372
Db 214 SVCEGSPSSSGSGRGLSLSTICFGSPCTSTGCGTLST-----SVSGGSSSTISANF 267
QY 373 GLIKRLGMMIESLEIDPTDSNMHLYGTGWTIFGCHDLTNDTRHNVSIOGLADGIEFVS 432
Db 268 G-----GTLSTSLICFDSPST---GAG--FGG-----ALNTSASFGSV 300
QY 433 QDLASAPGSELLAAVGDNGFTFASRNDLGTSPOFWMATPTMATSTSVDAANSKSVY 492
Db 301 LNTSTGFGAMSTSA---DFGTLSTSVCFGSGSTSVSFGS--ALNTNAGY--GGAVSTNT 355
QY 493 RVGNTAGTQVAISSDGCATWSIDVADTSMN-GGTVAVASD-GDTILMSTASSG--VQRS 548
Db 356 DFGTILSTSVCFGSPSTSAFGGALNTNMFCAVSTASFGAATSAFCFGAPITNP 415
QY 549 QFGSFAVSLSLAPAGAVIASD---KTNVSVFYAGSGST---FYSKDTGSSFTRGPK--- 599
Db 416 GFGGAFSTSAFGCALSTADDFGTPSNSIGFGAAPSTSVSFGCAHGTSLCFGGAPESTSL 475
QY 600 -LGSAGTIRDIAMHPTAGTLVYSTDVGI---RSPSGCTTPOVSPALNTYQIALGVG 655
Db 476 CFGSASNTNLCFGGPSTSAFCFGATSPSCDGPSTSTGFSFG-----NGLSTNAGFG 528
QY 656 SGNMNLVAFGTGSPGRLVYASDGSASMTD--IOGSGGFSIDSTKV--AGSGTAGO 710
Db 529 GGLNTSA-GFGGGLTSAFGSGGLSTSSGDDGLTSAAGGPGTSTGFGGLTSAAG- 586
QY 711 VYVGTCRGVIFYAGTY-----GGGTG-----GTS----- 755
Db 587 -FSGGLTSAFGGGLTSDGFGGGLTNASFGSTLTSTASFGGGLSTSDGFGSRPNASF 645
QY 736 -----SSMKOSSSTSSASSSTTLRASSVSTRASTVTSKRTSSAAGPTGSGVA 784
Db 646 DRGLSTILGSGSNTSTGTGEPSTSTGSSGPPSLVGFSGPSTGVCSCGSPSTSGFS 705
QY 785 GHYAOCGGIGMTGP 798
Db 706 GGPSTGAGFG-GGP 718

RESULT 9
```

```

: Sequence 2, Application US/08439818A
: US-08-439-818A-2
: Patent No. 5654145
: GENERAL INFORMATION:
:   APPLICANT: Fukuda, Michiko N.
:   TITLE OF INVENTION: Trophinin and Trophinin-Assisting
:   TITLE OF INVENTION: Proteins
:   NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Campbell and Flores
:   STREET: 4370 La Jolla Village Drive, Suite 700
:   CITY: San Diego
:   STATE: California
:   COUNTRY: USA
:   ZIP: 92122
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentln Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/439,818A
:     FILING DATE: 12-MAY-1995
:     CLASSIFICATION: 435
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: US 08/317,522
:       FILING DATE: 04-OCT-1994
:     ATTORNEY/AGENT INFORMATION:
:       NAME: Campbell, Cathryn A.
:       REGISTRATION NUMBER: 31,815
:       REFERENCE/DOCKET NUMBER: P-LA 1563
:       TELECOMMUNICATION INFORMATION:
:         TELEPHONE: (619) 535-9001
:         TELEFAX: (619) 535-8949
:     INFORMATION FOR SEQ ID NO: 2:
:       SEQUENCE CHARACTERISTICS:
:         LENGTH: 749 amino acids
:         TYPE: amino acid
:         TOPOLOGY: linear
:       MOLECULE TYPE: protein
: US-08-439-818A-2

```

```

0Y 493 RVGMTAGQVAVISSDGGATSNIDYADTSMN-GCTVAVYSAD-GDTILMSTASSG--VORS 548
Db 356 DEGGTLSTSVCFGSGSPSTSGAFGALNTMNSFGCAVSTASPSFGAVSTASCFSGARPTNP 415
0Y 549 QFGSGFASVSSLPGAVIASD--KKTNSVFYAGSGST---FVYSKDTGSSFTRGPK-- 599
Db 416 GFGGAFSTSGAFGALSTADAFGCTGPGTSPNSIGFCAAPSTSVSFGCAHOTSILCFGAPBSTSL 475
0Y 600 -LGSAGTTRDIAHPTTAGTLYIVTDVGI-F--RSTDSGTTFGQVSTALNTYQIALGVG 655
Db 476 CFGSASNTNLCEGSPPTSGACSFSGATSPSCDGPSTSTGFSFG-----NGLSTNAGFG 528
0Y 656 SGSMMNLIAFTGTCGARGALVACSGDGSAMD--IQSGQSGSIDSTKV---AGSGSTAGQ 710
Db 529 GGLMTASA-GFGCGGLGTSGAGFSGGLSTSSGFDGGLGTSGAGFGCGGPGCTGTGGGGLGTSGAG- 586
0Y 711 VYVGTNGHGVYVYAOGV-----GGGTG-----GTS----- 735
Db 587 -FSGGLGTSGAGFGGGLVYTDGFGGLTGNASFGSTLGTSGAGFSGGLSTSDGFGSRPNASF 645
0Y 736 -----SSTKSSSTSTSSASSSTTLRSSVSTTRASTVTSRSTSSAAGTGSVA 784
Db 646 DRGLSTIIIGFSSGSGNTSTGTFGEPTSTGTGSSGSSLVGEFGGPGTGVGFCSGPSTSGFS 705
0Y 785 GHVACCGGIGMTGP 798
Db 706 GGPSTGAGFGC-GCP 718

RESULT 10
US-08-751-965-2
Sequence 2, Application US/08751965
Patent No. 5858360
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/751,965
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 2252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-751-965-2

```

Query Match	Score	DB 2	Length
4.68	201	749	

Best Local Similarity 23.2%; Pred. NO. 4.4e-06;
Matches 184; Conservative 86; Mismatches 320

Matches 184; Conservative 86; Mismatches 320; Indels 204; Gaps 36;

QY	92	SNGLIITS--SDRATVSEFTLPEKVGNNMPGRAGRILAVDPANS-----NIIYFCA	142
Db	42	SRGASTRGESDR-ASISFNCAPSSCGF--SGGPITTCVAVBTA\$SFSNTASISGCT	98
QY	143	RSQGNLWKST-----DGVYFSKVSSEFATGCTITPPSPSNGYNSDKOGLMWTFD	1933
Db	99	LSTSSSFSMAASISFGCAHSTSTSFSS\$BASISFGCMPT\$AST\$SGCVSS\$FSCPL-----	153
QY	194	ST\$STTGCA\$TRIVGTAADNTIT\$AYVYSTNAG\$TWS\$AVPCQ\$PKYFPH\$K\$KLOPAEKALY	2533
Db	154	ST\$AT\$FSGG\$SSG\$GCT-----L\$T\$AG--F\$C\$V-----L\$	1822
QY	254	LTY\$GTC\$PYG\$G\$T\$G\$SV\$W\$R\$Y\$D\$IAG\$T\$K\$D\$IT\$PV\$G\$S\$D\$LY\$F\$G\$G\$LD\$Q\$K\$P\$G\$TLV\$V\$ASL	3133
Db	183	T\$T\$FSG\$A\$P\$T\$T\$Y\$F\$S-----S\$AL\$T\$G\$F\$G\$-----I\$L\$T	2133
QY	314	NS\$M\$P\$D\$Q\$L\$F\$R\$S\$T\$D\$G\$T\$T\$W\$P\$IM\$A\$M\$A\$S\$P\$T\$E\$T--Y\$Y\$S\$I\$P\$K\$A\$P\$W\$IK\$N\$E\$ID\$Y\$T\$E\$S\$P\$D	3722
Db	214	SV\$C\$G\$S\$P\$S\$S\$G\$G\$C\$T\$ST\$IC\$F\$G\$G\$S\$P\$C\$T\$G\$G\$T\$L\$T\$-----SV\$F\$G\$G\$S\$S\$T\$AN\$F	2672
QY	373	GL\$IK\$RL\$G\$M\$IS\$LE\$ID\$P\$D\$S\$N\$H\$M\$LY\$G\$T\$G\$M\$IT\$F\$G\$H\$D\$LT\$M\$D\$R\$H\$N\$V\$T\$O\$L\$A\$G\$IE\$E\$F\$V	4422
Db	268	G-----G\$T\$L\$T\$S\$IC\$F\$D\$S\$P\$T\$-----G\$G\$A\$-----F\$G\$-----A\$L\$N\$T\$S\$A\$F\$S\$V	3002
QY	433	QD\$L\$A\$P\$G\$S\$E\$L\$A\$A\$V\$D\$D\$N\$G\$F\$T\$F\$A\$S\$R\$N\$D\$G\$T\$S\$P\$Q\$Y\$M\$A\$T\$P\$T\$W\$A\$T\$S\$V\$D\$Y\$A\$G\$N\$S\$V\$Y\$V	4922
Db	301	L\$N\$T\$G\$T\$G\$C\$A\$M\$S\$T\$A\$-----D\$R\$G\$T\$L\$S\$T\$V\$C\$F\$G\$S\$P\$C\$T\$S\$V\$S\$F\$G\$S--A\$L\$N\$T\$N\$A\$G--G\$A\$V\$S\$T\$N\$T	3552
QY	493	R\$V\$N\$T\$A\$G\$T\$Q\$V\$A\$I\$S\$D\$G\$G\$A\$T\$W\$S\$I\$D\$Y\$A\$D\$T\$S\$M\$N--G\$G\$T\$A\$Y\$S\$A\$D--G\$D\$T\$IL\$M\$T\$A\$S\$G\$--V\$O\$R\$	5482
Db	356	D\$F\$G\$T\$L\$S\$T\$S\$V\$C\$F\$G\$S\$P\$S\$T\$S\$A\$G\$F\$G\$A\$L\$T\$M\$N\$A\$S\$F\$G\$C\$A\$V\$S\$T\$S\$F\$G\$A\$V\$S\$T\$S\$A\$C\$F\$G\$A\$P\$T\$N\$P	4152
QY	549	O\$F\$G\$S\$P\$A\$S\$V\$S\$L\$P\$A\$G\$A\$V\$I\$A\$S\$D\$-----K\$T\$N\$S\$V\$F\$Y\$A\$G\$S\$G\$T\$-----F\$Y\$S\$K\$T\$G\$S\$F\$T\$R\$G\$P\$-----	5592
Db	416	G\$F\$G\$A\$F\$S\$T\$S\$A\$G\$F\$G\$A\$L\$S\$T\$A\$D\$F\$G\$T\$P\$N\$S\$T\$G\$F\$G\$A\$P\$S\$T\$S\$V\$F\$G\$A\$H\$T\$S\$IC\$F\$G\$A\$B\$T\$S\$L	4752
QY	600	L\$G\$S\$A\$G\$T\$R\$D\$IA\$H\$P\$T\$A\$G\$T\$LY\$V\$S\$T\$D\$V\$G\$IF-----R\$S\$D\$S\$G\$T\$T\$F\$Q\$V\$S\$T\$A\$L\$N\$T\$Y\$Q\$IA\$LG\$V\$G	6552
Db	476	C\$F\$G\$S\$M\$T\$N\$T\$LC\$F\$G\$P\$P\$T\$S\$A\$C\$F\$G\$A\$T\$S\$P\$R\$C\$D\$G\$S\$T\$G\$S\$F\$G-----N\$G\$L\$S\$T\$N\$A\$G\$F\$G	5282
QY	656	S\$G\$S\$M\$N\$LY\$A\$B\$T\$G\$S\$G\$A\$R\$LY\$A\$S\$G\$S\$G\$A\$S\$M\$D\$--I\$Q\$S\$Q\$G\$C\$S\$ID\$S\$T\$K\$V-----A\$G\$S\$T\$A\$G\$O	7102
Db	529	G\$G\$L\$T\$S\$A--G\$F\$G\$G\$C\$G\$T\$S\$A\$G\$F\$G\$G\$C\$T\$S\$T\$S\$G\$C\$D\$G\$C\$G\$T\$S\$A\$G\$G\$G\$G\$P\$C\$T\$S\$T\$G\$F\$G\$G\$C\$G\$T\$S\$A\$G--	5862
QY	711	V\$Y\$E\$N\$G\$R\$V\$Y\$A\$G\$T\$V\$-----G\$G\$G\$T\$G-----G\$T\$S-----	7352
Db	587	F\$S\$G\$G\$L\$G\$T\$S\$A\$G\$F\$G\$G\$LY\$T\$S\$D\$G\$F\$G\$G\$C\$G\$T\$N\$A\$S\$F\$G\$T\$C\$T\$S\$A\$G\$F\$S\$G\$G\$S\$L\$S\$D\$G\$F\$G\$S\$R\$P\$A\$S\$F	6452
QY	736	-----S\$T\$K\$Q\$S\$S\$S\$T\$S\$A\$S\$S\$T\$L\$R\$S\$V\$S\$T\$T\$R\$A\$S\$T\$Y\$S\$T\$S\$S\$A\$A\$G\$P\$T\$G\$S\$V\$A	7842
Db	646	D\$R\$G\$L\$T\$T\$IG\$F\$G\$S\$G\$N\$T\$S\$T\$G\$F\$G\$B\$E\$S\$T\$S\$T\$G\$F\$S\$G\$P\$S\$S\$IV\$G\$F\$S\$G\$P\$T\$G\$C\$F\$G\$S\$P\$S\$T\$S\$G\$F\$S	7052
QY	785	G\$H\$Y\$A\$Q\$G\$C\$IG\$M\$T\$C\$P	7982
Db	706	G\$G\$P\$T\$G\$A\$G\$F\$G--G\$P	7182

CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/738,975

FILING DATE: herewith

CLASSIFICATION: 435
 PRIOR ADDITION DATA:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: IIS 08/439 818

APPLICATION NUMBER: US 08/439,818
FILING DATE: 05-Dec-1995

FILED DATE: 03 DEC 1999
ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn

REGISTRATION NUMBER: 31,

REFERENCE/DOCKET NUMBER: P-LA 2251

TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001

TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 2

SEQUENCE CHARACTERISTICS:

LENGTH: 749 amino acids

TYPE: amino acid
TOPOLOGY: 1 inner

TOPOLOGY: linear

MOLECULE TYPE: protein
8-738-975-2

1
2
3
4
5
6
7

every match	4.68;
-------------	-------

Query Match	4.6%	Score	201	DB 2	Length	749			
Best Local Similarity	23.2%	Pred. No.	4.4e-06						
Matches	184	Conservative	86	Mismatches	320	Indels	204	Gaps	36

```

0Y 92 SNAHITSS--SDRATWSPFNLPKVCGNMPGRACERLAVDPANS-----NIIYFGA 142
Db 42 SRGSTRAGSDDR-ASISFGNABSSSGGF--SGGPGITTCGABPSTSAFSENTIASISFGT 98
0Y 143 RSGNGLKST-----DGGVTFKSVSFFTAGTITPPSPDSNGYNSKOGILMWTFD 193
Db 99 LSTSSPSSAASISFGCAHSTSTFSSEASISFGCMCTASFSFGCVSSSFGPL----- 153
0Y 194 STSTTGATGASRIIVGADNINTASVYSTNAGTWSAIVPQGPCKYFPHKKLQPAEKALY 253
Db 154 STSTSFSGGASSGGGFT-----LSTTAG--FSGV-----LS 182
0Y 254 LTYSDGTGPRPDGTGTSWRYRDIAGGTHKDDTPVGSGLYFGFGLGLDLOKPGTLVVASL 313
Db 183 TSTSFSGAPTTSTVFS-----SALSTSGFGS-----ILST 213
0Y 314 NSMWPPDQLEFRSDSGTTWBPIDYAMASYPIET--VYISITSPKAPWIKNNTIDVTSBPSD 372
Db 214 SVCGGSGPSSSGSGFGTSLSTICGGSGPCTSGTGGTLST-----SVSGGSSSTSANF 267
0Y 373 GLIKRLAMTIESLEIDPTDSNHMLYTGMTIFGGHDLTNDTRHNVSIOGLADIEFVS 432
Db 268 C-----GTLSTSLCFDPSPT--GAG--FGS-----ALNTSASGCV 300
0Y 433 QDLASAPGSGELLAAVGDDNGFTFASHNDLGTSPQYMATPTMATSTSVYDAGNSYKSV 492
Db 301 LNTSTGFGGAMSTSA--DEFGTLSTSVCFGSPGSPCTSVSFGS--ALNTNAGY--GGAVSTNT 355
0Y 493 RVGNTACTOVAISSDGGATMSIDVAAPTSMN--GGTVAYSD--GDTILMSTASSG--VQRS 548
Db 356 DFGGTLSTSVCFGSPGSPSTSAFGGALTNWNSFGCAVSTSAFSGAVSTSAFCFGAPITNP 415
0Y 549 QFOGSPFASVSLPAGAVIASD---KRTNSVYFAGSGST---FYVSKDTGGSFTRGPK-- 599
Db 416 GFGGAFSTSAFGGALSTADPFGCTPENSIGFGAAPTSTVSFGGAHOTSILCFGAPPTSL 475
0Y 600 -LGSAGTIRIDIAHPHTAGTILVYSTDVGIF--RSTDSGTTFGGVSTALTNTVQIALGVG 655

```

Db 476 CPGSASNTNLCFGPSPSTACFSGATSPSPFCDDPSTSTGFSFG-----NGLSTNAGFC 528
Oy 656 SCSNMNLYAFGTGPGSARLYASGDSGAWTD--IOGSGFGSIDSITKV--AGSGTAGQ 710
Db 529 GGLNTSA-GRGGGLGTSAGSGSLSTSSGDFDGLGTSAGGCGPGSTGCGGLGTSAG- 586
Oy 711 VYVGTNGRGVFAQGYV-----GGGTG-----GTS----- 735
Db 587 -FSGGLGTSAGFGGLVTSDFGGLGTNASFSGSLTGTAGFSGLSTSDGFGSRPNASF 645
Oy 736 -----SSTKSSSTSSASSSTILRSSVSTRASTYTSSTSSAAGPTSGVA 784
Db 646 DRGLSTILIGFGSGSNSTGTGTGPTSTGTGSSGSPSSIVGFGSGPSTGVGCGSPSTSGFS 705
Oy 785 GHYAOCGIGTGP 798
Db 706 GGPSTGAGFC-GGP 718

RESULT 12
US-08-728-626-2
Sequence 2, Application US/08728626
Patent No. 5910451
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,626
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 1563
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-728-626-2

Query Match 4.6%; Score 201; DB 2; Length 749;
Best Local Similarity 23.2%; Pred. No. 4.4e-06;
Matches 184; Conservative 86; Mismatches 320; Indels 204; Gaps 36;

Oy 92 SNGAIIRS--SDRGATWSFNLFPKVGNNMPGRGAGERLAVDPANS-----NIIFYGA 142
Db 42 SRGASRTRAGSDR-ASISFNGAPSSSGGF--SGGPGITTFVAPSTASFSFNTASISISGCT 98
Oy 143 RSGNGIMKST-----DGVTFKSVSSFTATGYIIPDPSDNGYNSDKGLMWTFD 193

Db 99 LSTSSSFSSMAASISFGCAHSTSTFSFSSSEASISFGCMPTCSAFSGSVSSFSFGL----- 153
Oy 194 STSTSTTGATSRLEFVGTADNITASVYSTNAGSTWASVPOPKRYPHKAKLOPAEKALY 253
Db 154 STSATFSGGASSGFGGT-----LSTTAG--FSGV-----LS 182
Oy 254 LTVSDGTGPDYDGLGWSRWYDIAGTWKIDITPVSGSDLYFGFGLDLQKPGTLVVASL 313
Db 183 TSTFSGSAPTTSTVPS-----SALSTGFGG-----ILST 213
Oy 314 NSWMPDAQLFRSDSGTWSPIWAMASYPTET-YYYSISTPKAPWKNFIDVTSFSPSD 372
Db 214 SVCFGSSPSSSGSFGGTLSTICFGGSPCTSTGFGTLST-----SVFSGSSSTISANF 267
Oy 373 GLIKRLGMIESLEIDPTDSNHWLYGTGMIFGCHDLTMDTHNVSIOGLAGIEEFVY 432
Db 268 G-----GTLSSTICFPGSPST--GAG--FSG-----ALNTSASFGSV 300
Oy 433 QDLASAPCGSELLAAVGDNDGFTFASRNDLGTSPQTVWATPTWATSTVDYAGNSYKSVY 492
Db 301 LNTSTGFGAMSTSA---DFGGLTSTSVCFGSGSPGTSVSTGS-ALNTNAGY-GGAVSTNT 355
Oy 493 RVGNTAGTOYAISSDGAATWSIDYADTSNN-GGTVAYSAD-GDTILMSTASSG--VQRS 548
Db 356 DFGGLTSTSVCFGSGSPSTSAFGGALNTNASFCAVSTASFGAVSTACFSGAPITNP 415
Oy 549 QFQGSFASVSLPAGAVIAD---KTNVSYFVAGSGST---FVSKDGTGSFTRGK--- 599
Db 416 GFGGARFSTSAFGGALSTADDFGTPSNISGFGAAPTSTVSFSGAGTSLCFGGAPOSTSL 475
Oy 600 -LTSAGTIRIDIAHPTTAGLTVSTDVGR---RSTDSGTTFGOVSTALNTQIALGVG 655
Db 476 CPGSASNTNLCFGPSPSTACFSGATSPSPFCDDPSTSTGFSFG-----NGLSTNAGFC 528
Oy 656 SCSNMNLYAFGTGPGSARLYASGDSGAWTD--IOGSGFGSIDSITKV--AGSGTAGQ 710
Db 529 GGLNTSA-GRGGGLGTSAGSGSLSTSSGDFDGLGTSAGGCGPGSTGCGGLGTSAG- 586
Oy 711 VYVGTNGRGVFAQGYV-----GGGTG-----GTS----- 735
Db 587 -FSGGLGTSAGFGGLVTSDFGGLGTNASFSGSLTGTAGFSGLSTSDGFGSRPNASF 645
Oy 736 -----SSTKSSSTSSASSSTILRSSVSTRASTYTSSTSSAAGPTSGVA 784
Db 646 DRGLSTILIGFGSGSNSTGTGTGPTSTGTGSSGSPSSIVGFGSGPSTGVGCGSPSTSGFS 705
Oy 785 GHYAOCGIGTGP 798
Db 706 GGPSTGAGFC-GGP 718

RESULT 13
US-08-808-5994-2
Sequence 2, Application US/088085994
Patent No. 6111089
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin, Trophinin-Assisting
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808,599A
FILING DATE: 28-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 2256
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-9949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-808-599A-2

```

```

Query Match      4.6% Score 201, DB 3, Length 749;
Best Local Similarity 23.2%; Pred. No. 4,4e-06;
Matches 184; Conservative 86; Mismatches 320; Indels 204; Caps 36;

```

```

QY 92 SNAIIIR--SDRGATMSTFNLPRKVGNMGRGACERLAVDPANS-----NIIFGA 142
DB 42 SRGASTAGSDR-ASISFCAPSSSGF--SGPGITFGVAPSTASFSNTASISFGT 98
QY 143 RSGNGLMKST-----DGVTFKVSSEPTATGYIPPSDSNGYNSDKGLMWTFD 193
DB 99 LSTSSSSASASISFGCAHSTSTSFSEASISFGMPCMTASFGVSSSGFL----- 153
QY 194 STSTTGATSRITVGTADNITASVYTNAGSTWASVPGOPGKFPKAKLQPAEKALY 253
DB 154 STSTFSGGASGSGFGT-----LSTAG--FSGV-----LS 182
QY 254 LTVSDGTGPDGTGLSVWRDIAGCTWKDITPVSGSDLYFGGGLGLDKPGTLVVASL 313
DB 183 TSTSGSAPTTSTYF-----SALSTSTGFGG-----ILST 213
QY 314 NSMWPADQLFRSTDCTTWSPTWAMASYPTET-VYYSISTPKAWIKNFIDYTESPSD 372
DB 214 SVCGGSGPSSSGRGLSTSLICFGGSPCTSTGCGTLST-----SVSGGSSSTANF 267
QY 373 GLIKRLWMIESLEIDPTDSNHLIYGTGMITFGCHDLTMDTRHNVSIOSLADGIEEFV 432
DB 268 G-----GTLSTSLICFDSGSPST---GAG---FGG-----ALNTSASFGSV 300
QY 433 ODLASAPGSGELAAVDDNGFTFASRNDLGTSPQVWATPTMATSTSYDAGNSVSKV 492
DB 301 LNTSTGCGANSTSA---DTGGLTSTVCGGSGPSTVSFGS--ALNTNAGT--GGAVSTNT 355
QY 493 RVGNTACTOVAISSDGGATWSIDYADTSMN-GGTVAVSAD-GDTIIMSTASSG--VQRS 548
DB 356 DFGGTLSTVCGFGSPSTSGFGLALNTNMSFGCAVSTASFSGAVSTGACFSAPITNP 415
QY 549 QFGGSEFASVSLPAGAVIASD---KKTNSVYAGSGST---FYVSKDTGSEFTNGPK--- 599
DB 416 GFGGAFSTSGAGFGALSTADFCGTPSNSISICFGAAPTSTVSFGCAHGTSLICFGCAPSTSL 475
QY 600 -LGSAGTIRIOAHPTAGTLIVSTDGIF--RSTDSGTTFGGVSTALNTNYIALGVG 655
DB 476 CFGGASATNLCFGGPSTSAFCGATSPSFCDDGSTSTGFSFG-----NGLSTNAGFG 528
QY 656 SGNMNLVATCTGPGSARLYASCDGASWTD--TGGSGGFSIDSTKY--AGSGSTAGQ 710
DB 529 GGLNTSA-GTGGGLTSGAGSFGGLSTSSGFDGGLGTSGAGCGGCGTGTGCGGGLGTSGAG- 586

```

```

QY 711 VYVGTNGRVFYAGTV-----GGTG-----GTS----- 735
DB 587 -FSGGLCTSAGFGGLVTSDFGGLCTNMSFGSLTGTSGAGSGLSTDPGFSRPNASF 645
QY 736 -----SSTKSSSTSSASSTTLRSSVSTTRASTYTSSTSSAAGPTGSGVA 784
DB 646 DRGLSTIIGFGSGSNTSTGFTGERSTSTGFSSSPSIVGSGPSTGVGCSGFS7SGFS 705
QY 785 GHVAOCGIGMTGP 798
DB 706 GGPSTGAGFG-GGP 718

```

```

RESULT 14
US-08-676-166A-3
Sequence 3, Application US/08676166A
Patent No. 5955270
GENERAL INFORMATION:
APPLICANT: Radford, Alan
TITLE OF INVENTION: EXPLOITATION OF THE CELLULOSE COMPLEX OF
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,166A
FILING DATE: 15-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1321-1-002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: H. grisea
US-08-676-166A-3

```

```

Query Match      4.6% Score 200, DB 2, Length 525;
Best Local Similarity 19.7%; Pred. No. 3,2e-06;
Matches 149; Conservative 59; Mismatches 199; Indels 348; Caps 35;
QY 149 WKS-TDGVFESKVSSTFATGTYIPDPSDN-----GNSDKOGLMVT----- 191
DB 34 WKKTAGGCGCTVQASITL-----DSNWRVTHQVSGSTNCTYTNKKMDTSLCTDAKS 84
QY 192 -----FDSPS---STTGATSRITVGTADNITASVYTNAGSTWASVPGOPGKY 238
DB 85 CAQNCVADGADYTYSTVGTITNGDLSLKFV-----TKGY-STNVGSRFLMDGED--- 134
QY 239 FPKAKLQPAEKALYLYSDGTGTYDGLTSVWRDIAGTMDITVYS---GSDLYF-- 293

```


Wed May 7 14:15:34 2003

us-10-026-994-2.rai

Page 12

Search completed: April 26, 2003, 13:26:17
Job time : 33.4098 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:25:17 ; Search time 32.2509 Seconds
(without alignments)
2032.383 Million cell updates/sec

Title: US-10-026-994-2
Perfect score: 4343
Sequence: 1 AFSMKNVXKLGSGGFVPGII.....TCVAPVYCKQNDYYQCV 818

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications-AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	738	17.0	555	US-09-927-827-47	Sequence 47, Appl
2	225	5.2	1621	US-10-185-990-10	Sequence 10, Appl
3	225	5.2	1626	US-10-185-990-11	Sequence 11, Appl
4	222.5	5.1	2344	US-09-815-242-12713	Sequence 12713, A
5	211	4.9	1463	US-09-971-536-69	Sequence 69, Appl
6	209.5	4.8	596	US-10-063-547-100	Sequence 100, Appl
7	209.5	4.8	596	US-10-174-590-310	Sequence 310, Appl
8	209.5	4.8	596	US-10-176-758-310	Sequence 310, Appl
9	209.5	4.8	596	US-10-063-616-100	Sequence 100, Appl
10	209.5	4.8	596	US-10-175-737-310	Sequence 310, Appl
11	209.5	4.8	596	US-10-063-502-100	Sequence 100, Appl
12	209.5	4.8	596	US-10-173-706-310	Sequence 310, Appl
13	209.5	4.8	596	US-10-175-738-310	Sequence 310, Appl
14	209.5	4.8	596	US-10-176-752-310	Sequence 310, Appl
15	209.5	4.8	596	US-10-176-482-310	Sequence 310, Appl
16	209.5	4.8	596	US-10-176-757-310	Sequence 310, Appl
17	209.5	4.8	596	US-10-176-913-310	Sequence 310, Appl
18	209.5	4.8	596	US-10-180-552-310	Sequence 310, Appl
19	209.5	4.8	596	US-10-180-557-310	Sequence 310, Appl

20	209.5	4.8	596	US-10-173-700-310	Sequence 310, Appl
21	209.5	4.8	596	US-10-174-572-310	Sequence 310, Appl
22	209.5	4.8	596	US-10-174-579-310	Sequence 310, Appl
23	209.5	4.8	596	US-10-174-582-310	Sequence 310, Appl
24	209.5	4.8	596	US-10-174-588-310	Sequence 310, Appl
25	209.5	4.8	596	US-10-175-739-310	Sequence 310, Appl
26	209.5	4.8	596	US-10-175-740-310	Sequence 310, Appl
27	209.5	4.8	596	US-10-175-743-310	Sequence 310, Appl
28	209.5	4.8	596	US-10-176-488-310	Sequence 310, Appl
29	209.5	4.8	596	US-10-176-492-310	Sequence 310, Appl
30	209.5	4.8	596	US-10-176-747-310	Sequence 310, Appl
31	209.5	4.8	596	US-10-176-750-310	Sequence 310, Appl
32	209.5	4.8	596	US-10-176-985-310	Sequence 310, Appl
33	209.5	4.8	596	US-10-176-987-310	Sequence 310, Appl
34	209.5	4.8	596	US-10-176-991-310	Sequence 310, Appl
35	209.5	4.8	596	US-10-176-992-310	Sequence 310, Appl
36	209.5	4.8	596	US-10-176-993-310	Sequence 310, Appl
37	209.5	4.8	596	US-10-184-658-310	Sequence 310, Appl
38	209.5	4.8	596	US-10-173-695-310	Sequence 310, Appl
39	209.5	4.8	596	US-10-173-697-310	Sequence 310, Appl
40	209.5	4.8	596	US-10-173-705-310	Sequence 310, Appl
41	209.5	4.8	596	US-10-174-576-310	Sequence 310, Appl
42	209.5	4.8	596	US-10-174-585-310	Sequence 310, Appl
43	209.5	4.8	596	US-10-174-586-310	Sequence 310, Appl
44	209.5	4.8	596	US-10-175-747-310	Sequence 310, Appl
45	209.5	4.8	596	US-10-176-481-310	Sequence 310, Appl

ALIGNMENTS

RESULT 1
US-09-927-827-47
Sequence 47, Application US/09927827
Publication NO. US20030036176A1
GENERAL INFORMATION:
APPLICANT: Bower, Stanley G.
APPLICANT: Ramseier, Thomas M.
TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
FILE REFERENCE: 38-10(15824)B
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: US 60/279,493
PRIOR FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 69
SEQ ID NO 47
LENGTH: 555
TYPE: PRT
ORGANISM: Xanthomonas campestris
US-09-927-827-47

QY	2	FSMKNVKLGSGGFVPGIIFPKTKGVAVRTDGGIYRLNAD-DSWTAVTDGI-ADNAG	59
DB	78	YKMSVSAI-GGGGFTVGVLFHPAERGLAVARTDVGARMDAQAQQTALTDWIGADD--	134
QY	60	WHMKGIDAVLPODDQKVVYAAGVMTNSWDPSNGAIIRSSDRGATWSTNLPKVCNM	119
DB	135	WNLMGIDAFVADPADADALYLAAGTYMHE-RAGTAVALRSFNRRTFERADLDFKLGNO	193
QY	120	PGRGAGRLVDPANSIITIFGARGSGIKSTDGCTFEKVSFTMTGTYIYDP-----S	175
DB	194	LGRNGERLAVDPDGHVLLGSRDA-GLMRSDDRGHMAKVAVF-----DPAALGA	245
QY	176	DSNGYNSDKQ--GLMWTFTDSTSTTGATSRIFVGADNITFASVYVSTNAGSTWAVPG	233
DB	246	TARHHVREDAVGATFVFPFAAGSNGSPTRITYGVSTQST-SLYSEDAAGSMAPVAG	304
QY	234	QPCGYFPHAKLQPAEKA-----LYLTYSDDGTGTPYDGTGYSWRRYDIAGTKDITPV-	286

```

QY 111 LPFVVGMPGRGAGERLAVDPANSNIITYGAREGAGNMLKMTSDGCVFFSVSSSTAGTY 170
Dp 169 LPVPTTQNLQATATGSTLSDGNHSHRLIAGGSENTPACNHHDLLIAGYSTGTAGAGSTQTS 228
QY 171 I PPSDSNGYNYSKOGGLMMWTFEDSTSTGGATSRIFVGTADNITVASVYSTNAGSTWSA 230
Dp 229 GEDSSLTAGYGSTQ-----TAGGSNULTAGYGTGTAGSDSSLIAGYGSTQTSGGGSSL 282
QY 231 VPQGPQKTFPHKAKLOPAEKALITYT---SDGTCPPYDGTGLGSWRVDIAG-GTWKDLITPV 286
Dp 283 TAGY-----GSTQTAQEGSNULTAGYGTGTAGVDSSSL-----IAGYGS---TQT 323
QY 287 SGSS--LYFGEGGLDLOKRGTLVVASLMSWMPDADLPFSTDSGCTWSPIMMNASYPTE 344
Dp 324 SGSSSALTLAGG--STQTAQEGSNULTAGYGS-----GTAGSSSSLIAGYGS- 368
QY 345 TYYYSITPKAPWIKNNPIDTSESPDGLIKRLGMMIESLEIDPTDSNMLYGTGWTIF 404
Dp 369 -----TQTSGGD-----SSITAG-----YGSYTQTAQ 389
QY 405 GCHDLT-NMDTRHNVSIOS--LADGTEEFVSVDLASAPG-----GSELLAAVGD 451
Dp 390 EGSNULTAGYGTGTAGVDSSSLIAGYGSTQTSGGSSSALTAGYGTQTAQEGSNULTAGC-- 447
QY 452 NGFTFASRNDLGTSPTQYVWATPTWATSTSYDVAGNSVKSVMRVGNAGTAGOVAISSDQAT 511
Dp 448 -----STGTAGADSSLIAGYGSTQTSFGSSSLTAGYGT 481
QY 512 WSIDYADTSSNGTGV--ATVSADGDTILMTSTASSGVORSPQGSFASVSSLLPAGCAVIASD 569

```

Db 482 -----OTAREGSTLTAGYSGTGTAGADSSLIAGYSGTGTSG--SESSLTAGYSGTGT 531
Qy 570 KKTNSVFAGSGST-----FVMSDGTSSFTRGPK-----LG 601
Db 532 AQQGSVLTSGYSGTGTACAAASMLTGTGTGTAGHESFTIAGYSGTGTAGHKSILTAGY 591
Qy 602 SAGTIRD-----IAAHPTTAGTL-----YVSTDVGIFRS-----TDS 633
Db 592 STGTARDSDYLIAGYSGTGTAGSGSSLIAGYSGTGTASYSMLTAGYSGTGTAREHSDL 650
Qy 634 GTTGTQVSTALTNTQIALGVSGSGSNMMLYAFGTGPGSARLAYSASMTDIOGSG-- 691
Db 651 VTGTGSGTSSAGSSNLSL-LGYGS-----TGTAFKSLTAGYSGTGTADERSDLY 699
Qy 692 -GFGS-----IDSTFVAGSGTAGOVYGTNGRGVFAAG-----TVGGS--TGGS 735
Db 700 AGYGTSTAGYSSSLIAGYSGTGTGTAGYSGTGTADENSLLTGTGTSTAGYS 759
Qy 736 SSTQSSSTSSASSSTTLRSSVST---TRASTVTSRTSSAAGPTGSGVAGH 786
Db 760 SSLIAGYSGTGTAGYSGTGTLAGYSGTGTADERSDLYGTGTSTAGYASLIAGY 814

RESULT 4

US-09-815-242-12713
Sequence 12713, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 12713
LENGTH: 2344
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match 5.1%; Score 222.5; DB 10; Length 2344;
Best Local Similarity 20.8%; Pred. No. 9.4e-06;
Matches 203; Conservative 120; Mismatches 345; Indels 309; Gaps 45;

Qy 2 FSWKNVKG-----GGCG-----FVPGIIFHPKTKGVAARTDIGCL-----YRLNAD 44
Db 312 FSGK-VNIGNKRKEHGNGSDGIGFAFSPGVLGELNGAA---VGIGLSNMFGRKLDI- 366

Qy 45 DSWTAVTGTGDIADNCGHMMGIDAVALDPDDOKYVAAVGMVTNSDPN-----GATII 97
Db 45 DSWTAVTGTGDIADNCGHMMGIDAVALDPDDOKYVAAVGMVTNSDPN-----GATII 97

Db 367 -----YHN-----TSPNSSAKAKA-----DPSNVAGGAGFAGV 396
Qy 98 RSSDRGATSFNTLPFKYGMNPGRAGERLAVDPANS-----NIIEFG 141
Db 397 TTDSTGVASTYTS--SSRADN-----AAKLNVQDTNTTFDDFDININGDTKVMATKAG 448
Qy 142 ARSNGCL--WKSTDOGVYFSKYSSFTATGTYIIPDPSDNGYNSDKOGLMWTFFDSTSS-- 197
Db 449 QTWPNISDWIAKSGTINF--LSMTA-----STGATNLNQVQFGFEETESAV 496
Qy 198 -----TTG-----GATSRIFVTADN-----ITASVYSTAGSTWASV---- 231
Db 497 TQVRVDTTGTKDILPRTKYSNDOVV--TIDNQSALTAAGYVTSVDSYASTYNDT 554
Qy 232 -----PGQPKYFPKAKL-----OPAE-----KALYLTSD-GTG----- 261
Db 555 NKTVMATNAGOSVYIYFTDVAKAPYTVGNQITLVGKTNPIVLTITTDGTCVTNTVTGL 614
Qy 262 ----PYDGTLSV-----WRYDIAGGTWMDIPV--SG 288
Db 615 PSGLSYDSATNLSIOTPTKIGSTVTVSTDOANKSTTTFTINVDPTAPVPIDGKS 674
Qy 289 SDLPFGFGGLDLOKPGTLVVASLNSWMPDQLEFRSTDGTTSP-----IW 336
Db 675 SEVESPISPINATODNSGNAVNTVTGLPSGLTFDSTNMTISGPTNIGTITIVSTD 734
Qy 337 AMASYPTEYYYSISTPAPMIKNFIDVTSESPSDGLIKRLGMIESELEIDPDSNML 396
Db 735 ASGNKTTTTFKEVY-----RNSMSDSVSTSGT-----QOSQSVSTSKADQSAS 780
Qy 397 YGTGTTIFGHDLTNMDTRHNYSIQSLADGI-----EEFSV-----QDLA 436
Db 781 TSTSGSINTS--TSASYSKTSV--SLDSVSASKSLSTSESNVSSTSLVNSQSQS 836
Qy 437 SAPG-----SELAAVGDNGCTFASRND---LGTSPQTVATP 473
Db 837 SMSGSVSKSTSLSDPISNSSSTKESVSTSTSDSLSTSLSDSVMSST--GSLSKSQ 895
Qy 474 TWATSTVDYGNYSKSVRVNGTAGTOVALISDDGATMSIDYADTSMNGGTVAYSADG 533
Db 896 SLSTST--DSASTSQSVSDSTNSISTSESLSESSTSEISIS-----NSISNSVA 947
Qy 534 DTILMSTASSGVQRSQFQSGFASVSSLPAGAVIASDKKTNVYFVAGSGSTFYVSKDTGSS 593
Db 948 STSKLESQSTSLSTSDSKSMSTSESLSDSTSDSVSGSLVAGSQS--VSTSDS 1004
Qy 594 FTRGKLSAGTIRIDIAHPTTAGTLVYSTDVGIFRSTDSGTT--FGVSTALTNTYQIAL 652
Db 1005 MS-----TSEMISDSMSTSGSLAASDKSMVSMSSTSQSGSTSESLSDIST-- 1053
Qy 653 GVGSGSNMMLYAFGTGPGSAG--RLVNSGDSGASMTDIOGSGCGSI-----DS 698
Db 1054 ---SDSDSKSLSTSQSGSTSTSTSSVAMSQSQSTSGMSSTSQSDSTSTSTSESDS 1110
Qy 699 TKVAGSGTAGOVYGTNGRGVFAAGTGTGCTSSSTKQSSSTLARSV 758
Db 1111 TSDSKASATASESISQS-----VSTSGSVSTSLSTSESTSTSMDSSTSLSTSE 1165
Qy 759 VSTRASTVTSRTISSA 775
Db 1166 SDSTSDSTSDSISEA 1182

RESULT 5

US-09-971-536-69
Sequence 69, Application US/09971536
Patent No. US20020159976A1

GENERAL INFORMATION:

APPLICANT: Glenn, Matthew
APPLICANT: Havukkala, Ilkka
APPLICANT: Bloksberg, Leonard
APPLICANT: Dekker, James

```

: APPLICANT: Christenson, Anna
: APPLICANT: Holland, Ross
: APPLICANT: O'Toole, Paul
: APPLICANT: Reid, Julian
: APPLICANT: Coolbear, Timothy
: TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
: FILE REFERENCE: 1043c2
: CURRENT APPLICATION NUMBER: US/09/971,536
: PRIOR FILING DATE: 2001-10-02
: PRIOR APPLICATION NUMBER: U.S. No. US20020159976a1 09/634,238
: PRIOR FILING DATE: 2000-08-08
: PRIOR APPLICATION NUMBER: U.S. No. US20020159976a1 09/724,623
: PRIOR FILING DATE: 2000-11-28
: PRIOR APPLICATION NUMBER: PCT/NZ01/00160
: NUMBER OF SEQ ID NOS: 83
: SOFTWARE: fastseq for Windows Version 4.0
: SEQ ID NO: 69
: LENGTH: 1463
: TYPE: PRT
: ORGANISM: Lactobacillus rhamnosus
US-09-971-536-69

Query Match      4.9%: Score 211; DB 9; Length 1463:
Best Local Similarity 21.6%: Pred. No. 2.8e-05;
Matches 199; Conservative 115; Mismatches 384; Indels 222; Gaps 37;

QY 4 MNVNLKGG---GPFVPGIIFHFKTKGVA-----YARDIGGLVYLNMDSDSTANTDGI 54
DB 220 WBPVITKSKPKDGDIDSHSVFNVTYTGALMKGITTKAFIKGNKIQSGSVSLSDS 279
QY 55 ADNA-GMHWGIDAVLADPDQDKVYAAV---GMTNSDPSNGAIIRSDGATWSPFN 110
DB 280 TQTTGGLG-----QQNQISLEVLPGATY-----GTTGGIN 314
QY 111 LPPKGGN-MPRGAGERL-AVDPAANSITTFGASGCLMKSTDCGVTFKVSSTFATG 168
DB 315 VEYDDGSGTIVKGAIVNLTQRTDASKN-----ERGTNALITOGGNEFKDSTYILNK 369
QY 169 TYIPDSDSNGYNSDKGLMWTFFDSTSTGGATSRIFVGTADNTTAVVYSTNGSTM 228
DB 370 NALVKGDFPIYIED-GGULTYDKNATVSIY-GATGNIPY-----RIDGTGVNLNGS-- 421
QY 229 SAVPGOPGKYPFHAKLQPAEKALYLYTSD--GTGPDYDTLGSVWRYDIAGTWMKDITPV 286
DB 422 -----HMTITGCAPKLGVFINIKGTGFFVAASSTIDLANTGTGKSVNAI 469
QY 287 S-GSDLYEFG-----GGLG-LDLOKPGTLVVASLNSKWPDAOLFRTSDGTTWS 333
DB 470 NVANDQULSFADADANLITIDGTGEAHLIKYGD--DANINITYMPKSVLFKITDNDADS 527
QY 334 PIWMAVSYPTET-----YYYSITSPKAMIKNNFIDVSESPSD----- 372
DB 528 SLFKYSGTITLGOYKIIIPDDGNATGPKSAIYITIKGSSORTATVGEFTRAEQOSK 587
QY 373 -----GLIKRLGWMIES--LEIDP-TDSNHLVYGTGMITFGCHDLTNMOTRHVNSIOS 422
DB 588 ALADTFATKSLFEVYASDNFIKVPNTDETTL--TGKTTGAY-----VTISG 635
QY 423 LADGIEFVSVDLASAPGSELLAAVGDNGCTFPASRNDLGTSPQVWATPTWATISTVD 482
DB 636 L-KGIEPGL--TANSYDSTKYLVOADKGNMSY-----ELPTGV--SLPANASEFVLS 684
QY 483 YGNSYKSVVRGNTAGTOVAISSDGATWSIDYADTSMN-----GGVVAASAG 533
DB 685 SAGFIYKATVYINDAETPKOASSAGSLINANSADYTAOSAKATSAASDAASIASDA 744
QY 534 DTI-----LWSTASGVORSOF-----OGSFASVSSLPLAG 563
DB 745 QSLAGSHADNMELKSLASAEKOSQIALAASKSAASSAASALIVASASASASASAA 804
QY 564 AVIASDKTNSV-----FYAGSGSTFYYSKDTGSSFTRGPKLGSAGTIRDLAAHPTTACT 618
```

```

DB 805 AVSNADASANSAAAAYDSVASEASAAANDSSGVATASFAASAA-----AAMSALST 859
QY 619 LVYSTDVIGFRTSDGTTTGOVS-----TALT-----NTYQIALGVGSGS 658
DB 860 ADVAAVVAASDAAGSAAVAASAOOSDKNKQATATARSQALDLNLIKLTIDVASC 919
QY 659 NNMLVAFGTGPGARLYRAGSDGASWTDIQSGQFSGSIDSTKYVAGSGTAVGVYGTNCR 718
DB 920 SSSASAGCATATATATYASAASSASAEACSTAHQACSSADAVGSGSAAQIASTRASAA 979
QY 719 GVFFAAGTVGGGTGTSSSTKQSSSTSSASS-----TTLRSSVSTT 762
DB 980 SSYPKDGCIOSLASQASBAKASSAASAAATSAAVGFAASDASPOAKTAASADVASS 1039
QY 763 RASTVTSRSTSSAGPTGSG 782
DB 1040 AAST-ANSNAAASAAATKAG 1058

RESULT 6
US-10-063-547-100
: Sequence 100, Application US/10063547
: Publication No. US20020182638a1
: GENERAL INFORMATION:
: APPLICANT: Eaton, Dan L.
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Goddard, Audrey E.
: APPLICANT: Goddard, Paul J.
: APPLICANT: Grimaldi, Christopher J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3230R1C1
: CURRENT APPLICATION NUMBER: US/10/063,547
: PRIOR FILING DATE: 2002-05-02
: PRIOR APPLICATION removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 170
: SEQ ID NO: 100
: LENGTH: 596
: TYPE: PRT
: ORGANISM: Homo Sapien
US-10-063-547-100

Query Match      4.8%: Score 209.5; DB 9; Length 596:
Best Local Similarity 22.7%: Pred. No. 1.1e-05;
Matches 140; Conservative 77; Mismatches 228; Indels 171; Gaps 22;

QY 194 STSSTGATSRIFVGTADNTTASVYSTNAGSTWSAVPGOGKYPFHAKLQPAEKALY 253
DB 28 STSANTGSSVSSGASTATNAGSSV--TSSGVSTATISG----- 64
QY 254 LTYSDGTGPDYDTLGSVWRYDIAGTWMKDITPVSGDLYFGFGGLDLQKPGTLVVASL 313
DB 65 -----SVTSNGVSIYTNSEPHFTSSGIS-----TAT 91
QY 314 NSWMPDA-----QLFRTSDGTTWSPIMAMASYPTETETYYISITPKAPMKNFIDVTS 369
DB 92 NSEFTASSGISIATNSESSTSS--GASTATNS-----ESST 127
QY 370 PSDGLIKRLGWMIESLEIDPTDSNHLVYGTGMITFGCHDLTNMOTRHVNSIOSLADGIE 429
DB 128 PSSGA-----STVYNS--GSSVTSSGASTATNSESSTYSRSTATNSES 170
QY 430 FSVODLASAPGSELLAAVGDNGCTFPASRNDLGTSPQVWATPTWATISTVDVAGNSYK 489
DB 171 STLSSGASTATNSD-----SSTSSGASTATNSESST--TSSGASTATNSES 215
QY 490 SYV--RVGNFAGTOVAISSDGATWSIDYADTSMNGCTVAVSADGDTILWSTASSGVOR 547
```

```

Db 216 STVSSRSTATNSESSTSSGASTATNSESRTSSNAGTATNSES-----STSSGAST 269
QY 548 SOFOGFSVSVSLPAGAVIASDKTNSVFPAGSGSTFYVSKDQSSFTRGPKLGSAGTIR 607
Db 270 ATMSDS-----STVSSGASTATNSESSTSSGASTATNSESSTSSG-----311
QY 608 DIAHPTAGTIVSTDVGIFFRSTDSGTTFGQVSTALJNTYQIALGVSGSNMNLAFGT 667
Db 312 --ASTATNSDSSTSSGASTATNSESSTSSGIST-VTNSES-----ST 352
QY 668 GPGGALLVAGSDGAWTDQSGFGSJDSTVYVAGSGSTAGVYVGTNGRCVFAQGY 727
Db 353 PSSGANTATNSESSTSSGANTATNSESSTVSSGASTATNSESSTSSGANTATNSE 409
QY 728 GCGT-GCTSSSTKQSSSTSS-ASSSTLRSSVSTTRASTVT---SRTSSAAG-PTGS 781
Db 410 SSTSSGASTATNDSSTSSSEASTATNSESSTVS-SGISTVYNSESSTSSGANTATNS 468
QY 782 GVAGHYAOCGIGMTG 797
Db 469 GSSVTSAGSGTALTG 484

```

RESULT 7

```

US-10-174-590-310
; Sequence 310, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-310

```

Query Match 4.8%; Score 209.5; DB 9; Length 596;
 Best Local Similarity 22.7%; Pred. No. 1.1e-05;
 Matches 140; Conservative 77; Mismatches 228; Indels 171; Gaps 22;

```

QY 194 STSSTGATSRIFVGTADITATVYVSTNAGSTWAVPQPKYPPHAKKLQPAEKALY 253
Db 28 STSANTGSSVSSGASTATNSESSTV---TSSGVSTATISG-----64
QY 254 LTVSDGTGPYDGTGLGSVWRDYAGTWKDIPTVSGDLVFFGGLGLDLQPGTLVAVSL 313
Db 65 -----SSVTSNGVSIYTNSEFHTSSGIS-----TAT 91
QY 314 NSMWPPA-----QLFRSTDSGTTWSPITAMASVPTETIYVYISITPKAPWIKNNIDVTSES 369
Db 92 NSESTASSGISISTATNSESSTSS-----GASTATNS-----EST 127
QY 370 PSLGILKRLQMWIESLEIDPTDSNMHLVGTGMTIFGCHDLTNDTRHNVSIOGLADQIEE 429
Db 128 PSSGA-----STVYNS-----GSSVTSAGASTATNSESSTSSRASTATNSES 170
QY 430 FSVODIASAPGSELLAAVGDNDGFTFASRNDLGTSPQVWATPTWATSTSDYAGNSVY 489

```

```

Db 171 STLSGASTATNDS-----SSTSSGASTATNSESST--TSSGASTATNSES 215
QY 490 SVV--RVGNATAGQOVAISSDGCATWSIDYAADTSMNGCVYVAPDDITIMSTVASSGVQR 547
Db 216 STVSSRSTATNSESSTSSGASTATNSESRTSSNAGTATNSES-----STSSGAST 269
QY 548 SOFOGFSVSVSLPAGAVIASDKTNSVFPAGSGSTFYVSKDQSSFTRGPKLGSAGTIR 607
Db 270 ATMSDS-----STVSSGASTATNSESSTSSGASTATNSESSTSSG-----311
QY 608 DIAHPTAGTIVSTDVGIFFRSTDSGTTFGQVSTALJNTYQIALGVSGSNMNLAFGT 667
Db 312 --ASTATNSDSSTSSGASTATNSESSTSSGIST-VTNSES-----ST 352
QY 668 GPGGALLVAGSDGAWTDQSGFGSJDSTVYVAGSGSTAGVYVGTNGRCVFAQGY 727
Db 353 PSSGANTATNSESSTSSGANTATNSESSTVSSGASTATNSESSTSSGANTATNSE 409
QY 728 GCGT-GCTSSSTKQSSSTSS-ASSSTLRSSVSTTRASTVT---SRTSSAAG-PTGS 781
Db 410 SSTSSGASTATNDSSTSSSEASTATNSESSTVS-SGISTVYNSESSTSSGANTATNS 468
QY 782 GVAGHYAOCGIGMTG 797
Db 469 GSSVTSAGSGTALTG 484

```

RESULT 8

```

US-10-176-758-310
; Sequence 310, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-310

```

Query Match 4.8%; Score 209.5; DB 9; Length 596;
 Best Local Similarity 22.7%; Pred. No. 1.1e-05;
 Matches 140; Conservative 77; Mismatches 228; Indels 171; Gaps 22;

```

QY 194 STSSTGATSRIFVGTADITATVYVSTNAGSTWAVPQPKYPPHAKKLQPAEKALY 253
Db 28 STSANTGSSVSSGASTATNSESSTV---TSSGVSTATISG-----64
QY 254 LTVSDGTGPYDGTGLGSVWRDYAGTWKDIPTVSGDLVFFGGLGLDLQPGTLVAVSL 313
Db 65 -----SSVTSNGVSIYTNSEFHTSSGIS-----TAT 91
QY 314 NSMWPPA-----QLFRSTDSGTTWSPITAMASVPTETIYVYISITPKAPWIKNNIDVTSES 369
Db 92 NSESTASSGISISTATNSESSTSS-----GASTATNS-----EST 127

```

```

0Y 370 P$DOLIKRLGMIESLEIDPDSNHHMLYGTGMIJFEGHDLTNDPTRHNVISLOAIOEE 429
Db 128 PSSGA-----STVNS---G$SVTSSG$ATNSESSTVSSRA$TATNSES 170
0Y 430 F$VODLASAPGSEL$LA$VGDNDGFE$ASRNDLCT$BPOTWATPMTAT$T$VDYAG$SVK 489
Db 171 $TLSSG$ASTATNSD-----S$TSSG$ASTATNSESST--TSSG$ASTNSES 215
0Y 490 $V$--R$C$NTAGTOVAISSD$GATW$SIDYADP$T$M$NC$GTA$Y$A$D$GT$ILM$T$A$S$SVOR 547
Db 216 $TVSS$R$A$T$ATN$S$S$TSSG$A$T$ATN$S$S$T$T$S$G$A$T$N$S$S$-----$TSSG$AST 269
0Y 548 $Q$FG$F$A$S$V$S$LE$G$A$V$A$S$D$K$T$N$V$F$A$G$S$T$P$E$V$K$D$T$G$S$P$T$R$G$K$G$A$C$T$IR 607
Db 270 ATNSDS-----$TVSSG$ASTATNSES$S$TSSG$ASTATNSES$S$TSSG----- 311
0Y 608 DIA$H$P$T$A$G$T$A$G$T$V$D$V$G$T$R$S$D$G$T$T$Q$V$S$T$A$T$N$T$Y$Q$A$L$G$V$G$S$M$N$N$A$F$G$T 667
Db 312 --A$T$AT$N$S$D$S$T$T$S$G$A$T$N$S$S$T$V$S$G$IS$T--$T$N$S$S-----$T 352
0Y 668 G$P$G$A$R$L$V$A$S$D$G$A$S$M$D$IG$S$G$F$S$ID$T$K$V$A$G$S$T$A$G$O$V$Y$V$T$N$G$R$G$V$F$A$O$G$T$V 727
Db 353 P$S$G$A$N$T$A$T$N$S$S$--T$T$S$G$A$T$A$T$N$S$S$E$S$T$V$S$G$A$T$A$T$N$S$S$E$S$T$T$S$G$V$T$A$N$S$E 409
0Y 728 G$G$T--G$C$T$S$T$K$O$S$S$T$S--A$S$S$T$T$L$R$S$V$V$T$T$R$A$S$T$V$T--$S$R$T$S$A$G--P$T$G$ 791
Db 410 S$T$T$S$G$A$S$T$A$T$N$S$D$S$T$T$S$E$A$T$A$T$N$S$E$S$T$V$S--G$IS$T$V$T$N$S$E$S$T$T$S$G$A$N$T$A$T$N$S 468
0Y 782 G$V$A$G$H$Y$A$D$C$G$IG$M$T$G 797
Db 469 G$S$V$T$S$A$G$S$T$A$T$L$T$G 484

```

```

RESULT 9
US-10-063-616-100
: Sequence 100, Application US/10063616
: Publication No. US20030013853A1
: GENERAL INFORMATION:
: APPLICANT: Eaton, Dan L.
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Gerlitsen, Mary E.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, Christopher J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Watanabe, Collin K.
: APPLICANT: Wood, William I.
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P3230R1C1
: CURRENT APPLICATION NUMBER: US/10/063, 616
: CURRENT FILING DATE: 2002-05-03
: Prior Application removed - See file wrapper or Palm
: NUMBER OF SEQ ID NOS: 170
: SEQ ID NO 100
: LENGTH: 596
: TYPE: PRT
: ORGANISM: Homo Sapien
: OS-10-063-616-100

```

Query Match	4.88;	Score 209.5;	DB 9;	Length 596;
-Best Local Similarity	22.78;	Pred. No. 1.1e-05;		
-Matches 140; Conservative	77;	Mismatches 228;	Indels 171;	Gaps 22;

```

Oy 194 STSTSGATGRITFVGTADNTATVYVSTNGTWSVAPPOGKFFPHAKLDPQEKALY 253
    11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
Db 28 STANATGSSVTSAGSATATNGSSV---TSSCGVTAATISC-----64

Oy 224 LYSNGTGPYDGTGLGSWVRWYIACGTAKKOLITPVSGSLYFEGFGLGLDLPGLVYASL 313
    11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:
Db 65 -----SSVTSNGVSIYVNSEFHTTSSIS-----TAT 91

Oy 314 NSMWMDA---QLFRSDSGTTSPIWMAVSPTEETVYVSIISFPAKIMKINFIDVTSSES 369
    11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11: 11:

```

```

Db      92  NSESTASSGISATINSESSSTTS-----GASTATNS-----ESST 127
QY      370  PSDGLIKRLGMMIESLEIDPTDSNHMLYGTGMTIFGCHDLTNDTRHNVSIOGLADIEE 429
Db      128  PSSGA-----STVTS-----GASVTSAGASTATNSESSTVSSRASTATNSES 170
QY      430  FSVODLASAPGSGSELLAAGDDNGFTFASRDLCSTQYTWALPTMATSTSVYAGNSVK 489
Db      171  STLSGSGASTATNSP-----SSTSSGASATATNSESST-----TSSGASTATNSES 215
QY      490  SVV---RVGNTGTGTOVAISSDGGATWASIDYAADTSMNGCTAAVSADDDTLMTSPASSGYOR 547
Db      216  STVSSRASTATNSESSTTSSGASATATNSESRTTNSNGAGTATNSES-----STTSSGASIT 269
QY      548  SOFGGSEFASVSLPAGAVIASDKKTNSEVYAGSGSTFVYSKDTGSSFTRGPKLGASGTR 607
Db      270  ATNSDS-----STVSSGASTATNSESSTTSSGASATATNSESSTTSSG----- 311
QY      608  DIAHPHTAGCLVYSTDVGIFRSDSGTTCQGVSTALTNTYQALVAGSGSMNNLVAFCR 667
Db      312  --ASTATNSDSSTTSSGAGTATNSESSTVSSGIST-VTNS-----ST 352
QY      668  GPSCARLYASGDSGASWMTDIOGSGGFSIDSTKYAGSGSTAGOVYGTNGRGVFAQGY 727
Db      353  PSSGANTATNSES---TTSGATATATNSESSTVSSGASATATNSESSTTSSGVSTATNSE 409
QY      728  GGGT-GGTSSSTKOSSSSTSS-ASSSTLLNBSVVSTTRASTVT---SSRTSSAG-PTGS 781
Db      410  SSTTSSGASTATNSDSSTVTSSEASTATATNSESSTVTS-SGITVTNSESSTTSSGANTATNS 468
QY      782  GVAGHYAQCGGIGMTG 797
Db      469  GSSVTSAGSGTAAALTG 484

```

```

US-10-175-737-310 RESULT 10
; Sequence 310, Application US/10175737
; Publication No. US2003003153A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Collin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-310

```

Query Match	4.8%	Score 209.5;	DB 9,	Length 596;
Best Local Similarity	22.7%	Pred. No. 1.1e-05;		
Matches 140; Conservative	77;	Mismatches 228;	Indels 171;	Gaps 22;

```
Qy      194 STSSTTGATSRIFEVGTADNITASVYVSTNGSGTWSAVPGQPKFYFPHAKLQPAEKALY   23
        ||| : | : | | | : || |:: : : |
Db      28 STSANIGSSVSISSGASTATNSGSSV--TSSGVSTATISG-----64
```

```

OY 254 LTVSDGTGPYDGLTGLSVWRIDYAGTWMKDTIPVSGSDLYFGFGLGLDLQPKGLTVASL 313
DB 65 -----SSVTNSGVSIYTNSEFHTSSGIS-----TAT 91
OY 314 NSWMPDA-----QLEFRSDGTTWSPIMAMASYPTETYYYSISIPKAPWKNNFIDVTSSES 369
DB 92 NSEFSTASSGISATINSESTTSS-----GASTATNS-----ESSST 127
OY 370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGCHDLTNMDTRHNVSIOIADGIEE 429
DB 128 PSSGA-----STVTNS-----GSSVTSSGASTATINSESTTSSRASTATINSES 170
OY 430 FSVYDLASAPGSELLAAVODDNGFTFASRNDLGTSPQYMATPTWATSTVDYAGNSYK 489
DB 171 STTSSGASTATNSD-----SSTTSSGASTATINSEST--TSSGASTATINSES 215
OY 490 SVY--RVGNMTAGTQVAISSDGCATWSIDYAADTSMNGGVAYASADDDTILMSTASSGVOR 547
DB 216 STVSSASTATINSESTTSSGASTATINSESRITSNAGATATINSES-----STTSSGAST 269
OY 548 SDFGGSFASVSLPAGAVIASDKTNSVFYAGSGSTFVYSGDTSFTRGPKLSAGTIR 607
DB 270 ATNSDS-----STVSSGASTATINSESTTSSGASTATINSESTTSSG-----311
OY 608 DIAHPTAGTLYSTVDGIFRSTDSGTTFGQVSTALNTYQIALGVSGSNMNLAFGT 667
DB 312 --ASTATNSDSSSTSSGAGATATINSESTVSSGIST--VTNSES-----ST 352
OY 668 GPGGALLYASGDGASMTDIOGSGFGSIDSTVYAGSGSTAGOVYVGTNGRVFYAGTIV 727
DB 353 PSSGANTATINSESS-----TSSGANTATINSESTVSSGASTATINSESTTSSGASTATINSE 409
OY 728 GGGT--GGTSSSTKQSSSTSS--ASSSTTLRSSVSTTRASTVT---SRTSSAAG--PTGS 781
DB 410 SSTTSSGASTATNSDSSSTTSSSEASTATINSESTVSS--SGISTVYNSESTTSSGANTATNS 468
OY 782 GVAGHYAQCGLGWTG 797
DB 469 GSSVTSSGSGTALTG 484

```

```

RESULT 11
US-10-063-502-100
; Sequence 100, Application US/10063502
; Publication No. US20030023042A1
; GENERAL INFORMATION:
; APPLICANT: Baton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerltisen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,502
; PRIORITY FILING DATE: 2002-05-01
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 100
; TYPE: PRT
; LENGTH: 596
; ORGANISM: Homo Sapien
US-10-063-502-100

```

```

Query Match 4.8%; Score 209.5; DB 9; Length 596;
Best Local Similarity 22.7%; Pred. No. 1.1e-05;
Matches 140; Conservative 77; Mismatches 228; Indels 171; Gaps 22;
OY 194 STSSTTGATSRIFVGTADNITASYVSTNAGSTWASVPGQPKRPYHPHAKLOPAEKALY 253

```

```

DB 28 STSANTGSSVISSGASTATINSCSV---TSSGASTATISG-----64
OY 254 LTVSDGTGPYDGLTGLSVWRIDYAGTWMKDTIPVSGSDLYFGFGLGLDLQPKGLTVASL 313
DB 65 -----SSVTNSGVSIYTNSEFHTSSGIS-----TAT 91
OY 314 NSWMPDA-----QLEFRSDGTTWSPIMAMASYPTETYYYSISIPKAPWKNNFIDVTSSES 369
DB 92 NSEFSTASSGISATINSESTTSS-----GASTATNS-----ESSST 127
OY 370 PSDGLIKRLGWMIESLEIDPTDSNHWLYGTGMTIFGCHDLTNMDTRHNVSIOIADGIEE 429
DB 128 PSSGA-----STVTNS-----GSSVTSSGASTATINSESTTSSRASTATINSES 170
OY 430 FSVYDLASAPGSELLAAVODDNGFTFASRNDLGTSPQYMATPTWATSTVDYAGNSYK 489
DB 171 STTSSGASTATNSD-----SSTTSSGASTATINSEST--TSSGASTATINSES 215
OY 490 SVY--RVGNMTAGTQVAISSDGCATWSIDYAADTSMNGGVAYASADDDTILMSTASSGVOR 547
DB 216 STVSSASTATINSESTTSSGASTATINSESRITSNAGATATINSES-----STTSSGAST 269
OY 548 SDFGGSFASVSLPAGAVIASDKTNSVFYAGSGSTFVYSGDTSFTRGPKLSAGTIR 607
DB 270 ATNSDS-----STVSSGASTATINSESTTSSGASTATINSESTTSSG-----311
OY 608 DIAHPTAGTLYSTVDGIFRSTDSGTTFGQVSTALNTYQIALGVSGSNMNLAFGT 667
DB 312 --ASTATNSDSSSTSSGAGATATINSESTVSSGIST--VTNSES-----ST 352
OY 668 GPGGALLYASGDGASMTDIOGSGFGSIDSTVYAGSGSTAGOVYVGTNGRVFYAGTIV 727
DB 353 PSSGANTATINSESS-----TSSGANTATINSESTVSSGASTATINSESTTSSGASTATINSE 409
OY 728 GGGT--GGTSSSTKQSSSTSS--ASSSTTLRSSVSTTRASTVT---SRTSSAAG--PTGS 781
DB 410 SSTTSSGASTATNSDSSSTTSSSEASTATINSESTVSS--SGISTVYNSESTTSSGANTATNS 468
OY 782 GVAGHYAQCGLGWTG 797
DB 469 GSSVTSSGSGTALTG 484

```

```

RESULT 12
US-10-173-706-310
; Sequence 310, Application US/10173706
; Publication No. US20030022293A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; PRIORITY FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; TYPE: PRT
; LENGTH: 596
; ORGANISM: Homo Sapien
US-10-173-706-310

```



```
Query Match Similarity      4.8%; Score 209.5; DB 9; Length 596.  
Match Local Similarity    22.7%; Pred No. 1.1e-05;  
Matches 140; Conservative   77; Mismatches 228; Indels 171; Gaps 22.
```

ID	Name
OY	STSSSTTGATSKRIEFGTADNITASYVSVINAGCTWSAVPGCGKYFPHKAKLOPAEKALY 253
Dd	STSANTGSSVISSGASTATNSGSVV---TSQGVSTATISG----- 64
OY	LTYSGGTGPYDGTLCASVMRYDIAGCTMKWDIPYPSCSDLYFEGGLGLDLKRGTLVASL 313
Dd	65 -----SSTSMNGCVSIVTNSEPHHTSSGIS-----TAT 91
OY	314 NSMWPD-----QLFRSTDGCTWSPIMAWASPTEHYYSISTPRAPWKXNFIDVTTSES 369
Dd	92 NSEFTASSGSIATATNESSTISS---GASTATNS-----ESST 127
OY	370 PSDGLIKRLGMIESLEIETPDSNHMLYCTGMTTCIGHDLTMDTRHNVSLDGIEE 429
Dd	128 PSSGA-----SYTVNS---GSSYTSSGGASTATNEESTSYSRSRSTATNSES 170
OY	430 FSVODLASAPGESSELLAAVGDNDNGTFPASRDNLGTSPOTWTATPTWAINTSVDYGNVSX 489
Dd	171 STLSGCASTATMSD-----SSTSSGCASTATNSESST--TSSGCASTATNSES 215
OY	490 SVV---RVENACTOYAIVSIDSGCATMSIDYAATSNNMGCYVAASAOGDTILMTASSGVOR 547
Dd	216 STVSRRASTATINESSTSSGSTATNTSESRITSNCAGATATNES----STSSGAST 269
OY	548 SOFOGSPFASVSLLPGAVALIASDKKTNSVFAYAGSGSTFYVKDPGTGSSFRRGPLKSAGRIR 607
Dd	270 ATNSQS-----STVSSGASATATNESESTSSSGASTATNESSSTSSG----- 311
OY	608 DIAHPHTTAGTILYSTVDVGIFIRSTDGGTFEGGYSTALTNYOLALGVSGSNMWNLYAFGT 667
Dd	312 --AATATNSDSSETSSGACTATNSESSTVSGISR-VTNSES-----ST 352
OY	668 GPSGARLYASGDSGSAWMDDIOSGOFGSIDSTRKKVAGSGGTACQOVYGTVGRGFVAOCGIV 727
Dd	353 PESSGANJATANSESS---TTSSCANMATATNESESTVSSGASTATNESESTSSSTATNSE 409
OY	728 GGCT-GCCTSSTPKOSSSTPS-ASSSTVLRSSVTVTTRAITYT---SRTSSAAG-PGCS 781
Dd	410 SSTTSSGASTATANSOSSTTSSBSAATATNSESSTVS-SGIStVTNBESSTSSGCANTATINS 468
OY	782 GVAGHYAACGCCIGIWG 797
Dd	469 GSSVTSAGSGTAALRG 484

RESULT 13
US-10-175-738-310
Sequence 310, Application US/10175738
Publication No. US20030022294A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Gudowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
FILE REFERENCE: P3430RIC45
CURRENT APPLICATION NUMBER: US/10/175.738
PRIOR ART CITATION DATE: 2002-06-19
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 310

```

: LENGTH: 596
: TYPE: PRT
: ORGANISM: Homo Sapien
us-10-175-738-310

Query Match          4.8% Score 209.5; DB 9; Length 596;
Best Local Similarity 22.7%; Pred. No. 1.1e-05;
Matches 140; Conservative 77; Mismatches 226; Indels 171; Gaps 22;

OY 194 STSSTTGATGATSRIFVGTADNITASYVSTNMGSTWSAVPGCGKYPHKAKLQPAEKALY 253
||| : : : : : ||| : : : : :
Db 28 STSANTGSSVSSSGASTATNCGSSV--TSSGVSATATISG----- 64

OY 254 LTYSDGTGTPYDGTLCGSVMRYDIAGCTMKDITPYGSSDLYFGRGGGLDLQKRGTLVAVSL 313
||| : : : : : ||| : : : : :
Db 65 -----SSVTNCGVSLVTPNTEPHHTSSGSL--TAT 91

OY 314 NSMWPEA----QLFRSTDGTTWSPIMWMAVPTETYYYSISTPRAPWIKKNFIDVTS 369
||| : : : : : ||| : : : : :
Db 92 NSEFTASSGSLSIATNSESSTTS--GASTATNS-----EST 127

OY 370 PSDGLIKRLGWMIESLEIDPTDSNMHLGYCTGTTIGCHDLTNMDRHNVSIOSLADGIEE 429
||| : : : : : ||| : : : : :
Db 128 PSSGA-----STVTNS--GSSVTSSGASTATNESSESYVSRASRTATNES 170

OY 430 FSVODLASAPGSGELLAAVGDDNGFTFASRNDLGTSPOTWATPTWATSTSYDAVAGNSVK 489
||| : : : : : ||| : : : : :
Db 171 STLSSGASTATNSD-----SSTTSSGASTATNSESST--TSSGASTATNES 215

OY 490 SVV--RVGNTAGTQVAISDGGATWSIDYADTSMNNGTVAVSAAGDIIIMSTASGVQR 547
||| : : : : : ||| : : : : :
Db 216 STVSSRASTATNSESSTTSSGASTATNSESRTTSMGAGTATNES-----STTSSGAST 269

OY 548 SOFGSFFASVSSLPAGAVIASDPKATNSVFYAGSGSTFYVSKDTCGSSFTRGPKLGSAGTIR 607
||| : : : : : ||| : : : : :
Db 270 ATNDS-----STVSSGASTATNSESSTTSSGASTATNSESSTTSSG----- 311

OY 608 DIAHPHTAGTLVYSTDVGIFRSTDSGTFEGGVSTALTNTYQIALGVSGSNMNLVAFGT 667
||| : : : : : ||| : : : : :
Db 312 --ASTATNSDSSTTSSGAGTATNSESSTVSGISL--VTNSES-----ST 352

OY 668 GPSGARLIASGSGSASWMDIOGSGCFGSLDSTKKVAGSGSTAGQVYVGTNGRCVFYAGQTV 727.
||| : : : : : ||| : : : : :
Db 353 PSSGANTATNSES--TSSGANTATNSESSTVSSGASTATNSESSTTSSGASTATNSE 409

OY 728 GCGT-GCTSSSTXKOGSSSTSS--ASSSTJLRSSVVTTRASTVT---SRTSSAAG-PTGS 781
||| : : : : : ||| : : : : :
Db 410 SSTTSSGASTATNDSSTTSSSEASTATNSESSTVS-SGLSTVTNSESSTTSSGANTATINS 468

OY 782 GVAGHYAOCGGIGWTG 797
||| : : : : : ||| : : : : :
Db 469 GSSVTSAGSGTMAALTG 484

RESULT 14
US-10-175-752-310
: Sequence 310, Application us/10175752
: Publication No. US2003002295A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Chen, Jian
: APPLICANT: Desnoyers, Luc
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Smith, Victoria
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3430R1C60

```

```

:
:
:
: CURRENT APPLICATION NUMBER: US-10/175,752
:
: CURRENT FILING DATE: 2002-06-19
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 612
:
: SEQ ID NO 310
:
: LENGTH: 596
:
: TYPE: PRT
:
: ORGANISM: Homo Sapien
:
: US-10-175-752-310

```

Query Match	4.88;	Score 209.5;	DB 9;	Length 596;
Best Local Similarity	22.78;	Pred. No. 1.1e-05;		
Matches 140;	Conservative 77;	Mismatches 228;	Indels 171;	Gaps 22

```

OY 194 STSSGTGATRIEIVGADNATASVYSTNAGS7WSAVPQOPGKPEPHKAKLOPAEKALY 2533
Db 28 STSAN7GSSVSIASGAS7ATN8GSSV---TSSGVS7AT1SG-----64
OY 254 L7YSDG7GPPYDGTGLGS7WRNDYDAGCTMKDITPVSGSDLY7GFGGLDLOPGLTVYASL 3133
Db 65 -----SSV7TNGVS7IYNSEFH7TSSG1S-----TAT 91
OY 314 NSMMPDA-----QLFRSDGCT7TWSP1WAMAS7PE7E7YYS7IS7PKAWIKNFI7DV7SES 369
Db 92 NSER7TAS7SGIS7IATN8SS7TSS---GAS7ATN8-----ESST 127
OY 370 PSDGLIKRLG7M7IESLE7ID7PDS7NHMLYGTG7MT7FGGH7DL7TMD7RHNV7SIQSLAD7IEE 429
Db 128 PSSGA-----STV7TNS---GSSV7TSSGAS7ATN8SS7TVSSRAS7ATN8SES 170
OY 430 FSVYD7LAS7P7GSG7ELL7A7VDNCF7TAS7RNDLCTSPQ7YWA7T7YAT7STSV7DAGNSV7K 489
Db 171 STL7SGAS7ATN8D-----S7TSSGAS7ATN8SEST--TSSGAS7ATN8SES 215
OY 490 SVV--RVGNTAG7TOVAIS7DGCAT7MS7IDY7AD7T7MNG7CA7YAA7SADCD7T7LM7TAS7SGVOR 547
Db 216 STVS7RAS7ATN8SS7TSSGAS7ATN8SE7T7TNSGAG7ATN8SS-----S7TSSGAS7 269
OY 548 S7OFG7S7FAS7VSSL7PAG7AVIAS7DK7TNS7YFA7GSG7T7YVS7KDG7SST7RGP7KLGS7ACT7IR 607
Db 270 ATN8SDS---STVS7SGAS7ATN8SS7TSSGAS7ATN8SES7TSSG-----3111
OY 608 DIAHP7TAG7LY7STDV7G7IR7SDG7T7F7OVS7TAL7N7Q7IAL7GVGSG7NMN7LYAFGT 667
Db 312 --AS7ATN8SDS7TSSGAG7ATN8SS7TVSSG7IST-V7TNS8-----ST 352
OY 668 GPSCGAR7LYAS7DGS7AS7W7TDI7G7SGG7FGS7ID7TKV7AGSG7STAG7OY7V7G7N7RGV7FYA7G7TV 727
Db 353 PSSGANT7ATN8SS7S---T7TSSGANT7ATN8SS7TVSSGAS7ATN8SES7TSSGVS7ATN8SE 409
OY 728 GCG7T-GC7TSS7T7KOS7SS7TSS--ASS7T7LR7SV7T7RAS7V7T---SS7R7TSSAAG-PTGS 761
Db 410 SST7TSSGAS7ATN8SDS7TSS7SEAS7ATN8SES7TVS-SG7IST7VNS8S7TSSGANT7ATN8S 468
OY 782 GVAG7HYAQC7G7IG7MTG 797
Db 469 GSSV7TSSAGS7T7AAL7TG 484

RESULT 15
US-10-176-482-310
: Sequence 310, Application US/10176482
: Publication No. US20030022296A1
GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Chen, Jlan
: APPLICANT: Desnoyers, Luc
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Smith, Victoria
: APPLICANT: Watanabe, Colin K.

```

```

      APPLICANT : Wood, William I.
      APPLICANT : Zhang, Zemin
      TITLE OF INVENTION : SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
      TITLE OF INVENTION : ACIDS ENCODING THE SAME
      FILE REFERENCE : P3430R1C70
      CURRENT APPLICATION NUMBER : US/10/176,482
      CURRENT FILING DATE : 2002-06-20
      Prior Application removed - See file wrapper or Palm
      NUMBER OF SEQ ID NOS : 612
      SEQ ID NO 310
      LENGTH : 596
      TYPE : PRT
      ORGANISM : Homo Sapien
      us-10-176-482-310

      Query Match      4.8%  Score 209.5 : DB 9 : Length 596:
      Best Local Similarity 22.7%: Pred. No. 1.1e-05:
      Matches 140: Conservative 77: Mismatches 228: Indels 171: Gaps 22

      QY 194 STSSTTGGATSRIFPGADNITASYVSTNAGSTASAVBPQDKYFPHAKLQPAEKALY 253
          ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
      DB 28 STSANTGSSVSISSGASTATNSGSSV---TSSGVSTATISG----- 64

      QY 234 LTYSGCTGPYDCTLGSVMRYDIAGGTMDITVYSSGDLFFGEGGLDLQKPTLVASL 313
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB 65 -----:SVTSMNCVSLVTNSEFHTSSGIS-----TAT 91

      QY 314 NSMWPDAA---OLFRSTDGCTWSPPLWMAVSYPTETYYYSISTPKAPWIKNFIQVTSSES 369
          || : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB 92 NSEFTASSGISIAINSESTTSS---CASTATNS-----ESST 127

      QY 370 PSDGLIKRLGMWIELEIDPTDSNHLVYTGMTIFGHDLTWMDFRHNVSIQSLADGIEE 429
          ||| : : : : : : : : : : : : : : : : : : : : : : : :
      DB 128 PSSGA-----STVTNS---GSSVTSAGASTATNSESSTVSSRASTATNSES 170

      QY 430 FSVQDLASAPSGSELLAVGDDNGTFPARNDLGTSPOIVMAAPPTVATSTVDYAGNSVK 489
          : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB 171 STLSSGASTATNSD-----STTSSGASTATNSESST---TSSGASTATNSES 215

      QY 490 SVV--RVNGTACTOYVAISSDGATWSIDYAADTSMNGCTVAVSADODTLIMSTASSGVOR 547
          || : : : : : : : : : : : : : : : : : : : : : : : :
      DB 216 STVSSRASATATNSESSTTSSGASTATNSESPTTSMGACATATNSES-----STTSSGAST 269

      QY 548 SGFQSGSFASVSLPAGVAVIASDKRTNSVYVYASSGSEFYVSKTQGSFTTGPKLGSAGTR 607
          : : : : : : : : : : : : : : : : : : : : : : : : : :
      DB 270 ATNSDS---STVSSGASTATNSESSTTSSGASTATNSESSTTSSG----- 311

      QY 608 DIAHPPTAGTLVYTDVGIFRSTDSGCTFGGCVSTALPTVYQIALGVGSGSMNMLYAFGT 667
          ||| : : : : : : : : : : : : : : : : : : : : : : : :
      DB 312 --ASTATNSDSTTSSGACTATNSESSTVSSGIST-VTNSES-----ST 352

      QY 668 GPSGARLVASGDSGASWMDIQSGQGFGLIDSTKRVAGSGSTACOVYVGTNGRVFYAOGTV 727
          ||| : : : : : : : : : : : : : : : : : : : : : : : :
      DB 353 PSSGANTATNSES---TSSGANTATNSESSTVSSGASTATNSESSTTSSGCVSTATNSEE 409

      QY 728 GGGT--GCTSSSTKQSSSTSS--ASSSTTLRSSVSTTRASTYT---SSRTSSAG--PTGS 781
          || : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
      DB 410 SSTTSSGASTATNDSSTTSSSEASTATNSESSTVS--SGISTVYNSESSTTSSGANTATNS 468

      QY 782 GVAGHYAOCGGIGMTG 797
          || : : : : : ||
      DB 469 GSSVTSAGSGTALTG 484
  
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:33:03 ; Search time 54.7288 Seconds
(without alignments)
1436.866 Million cell updates/sec

Title: US-10-026-994-2

Perfect score: 4343
Sequence: 1 AFSWKNVKLGSGGFVPGII.....TQCVAPYVCKQNDYVYQCV 818

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR73:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2247	51.7	856	2 T00349	Avicelase III - As
2	1591.5	36.6	890	2 T35237	probably secreted
3	1500.5	34.5	839	2 D97013	probably secreted
4	899	20.7	707	2 F72933	hypothetical prote
5	313	7.2	2232	2 T34434	hypothetical prote
6	300.5	6.9	1275	2 T33369	hypothetical prote
7	267	6.1	1217	2 S52714	sericinB - silkwo
8	264.5	6.1	537	1 JU0150	cellulose 1,4-beta
9	263	6.1	13288	2 T03099	mucin, submaxillar
10	257.5	5.9	1567	2 S11672	ice nucleation pro
11	257	5.9	5281	2 F90696	hypothetical prote
12	255	5.9	5188	2 B85547	probable RTX fami
13	252.5	5.8	2468	2 A83412	hypothetical prote
14	243.5	5.6	1032	2 T34433	hypothetical prote
15	243.5	5.6	1258	2 JU0188	ice nucleation pro
16	241	5.5	1341	2 H98323	hypothetical prote
17	240	5.5	1254	2 AB3528	extracellular seri
18	236	5.4	1322	2 S07053	ice nucleation pro
19	230	5.3	1210	2 A25547	secreted acid phos
20	229.5	5.3	888	2 T46726	hypothetical prote
21	226.5	5.2	786	2 T16509	cellulose 1,4-beta
22	226	5.2	516	2 S42093	hypothetical prote
23	223.5	5.2	1200	1 SNPSO	extracellular seri
24	223	5.1	1374	2 AE3259	ice nucleation pro
25	221	5.1	805	2 T25795	hypothetical prote
26	220	5.1	2174	2 E95965	hypothetical glyco
27	219.5	5.1	3624	2 AD0835	large repetitive p
28	218	5.0	1649	2 C86822	hypothetical prote
29	217.5	5.0	605	2 T33913	hypothetical prote

30	217	5.0	1034	2 JC2143	ice nucleation act
31	216.5	5.0	827	2 JC4900	transferred entry
32	216.5	5.0	2271	2 F90073	hypothetical prote
33	216	5.0	1428	2 T08852	lustrin A - Califo
34	214.5	4.9	1910	2 AF0394	probable adhesin h
35	213	4.9	1904	2 T13256	tail-host specific
36	212	4.9	1797	2 T21889	hypothetical prote
37	212	4.9	1805	2 T21888	hypothetical prote
38	212	4.9	2204	2 A70524	probable PPE prote
39	211.5	4.9	1283	2 T39174	hypothetical Serin
40	209.5	4.8	1589	2 T42233	submaxillary mucin
41	209.5	4.8	2124	2 A28452	proteoglycan core
42	209	4.8	1131	2 T41144	hypothetical serin
43	207	4.8	1004	2 C82672	surface-exposed ou
44	207	4.8	1190	2 A82615	surface protein XP
45	206.5	4.8	948	2 T11678	hypothetical prote

ALIGNMENTS

RESULT 1

T00349

Avicelase III - Aspergillus aculeatus

C:Species: Aspergillus aculeatus

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Jul-1999

C:Accession: T00349

R:Araki, M.; Takada, G.; Kawaguchi, T.; Sumitani, J.

Submitted to the EMBL Data Library, June 1998

A:Description: Avicelase III from Aspergillus aculeatus.

A:Reference number: Z14141

A:Accession: T00349

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-856 (KAA>

A:Cross-references: EMBL:AB01511; NID:d1199887; PID:d1029971

C:Genetics:

A:Gene: avIII

C:Superfamily: fungal cellulose-binding domain homology

F:823-854/Domain: fungal cellulose-binding domain homology <FCB>

Query Match	51.7%	Score 2247	DB 2	Length 856
Best Local Similarity	52.4%	Pred. No. 1.7e-111		
Matches 442	Conservative 127	Mismatches 235	Indels 40	Gaps 14
QY	1	AFSWKNVKGSGGFVPGIIFHPKGVAAARPDIGLYLNDDSVATDGIADNAGW	60	
DB	25	AYTWKNVVTGGGGFTPGIVFNPSSAKGVAAYARTDIGAYRLNSDDTWTPLMDVWG	NDTW 83	
QY	61	HNMGIDAVLADPODDOKVYAVGMYTNSMDPSNGAIIRSSDRGATWSFTNLPKVGGMNP	120	
DB	84	HDMGIDALATDPYDTRVYAVGMYTNSMDPSNGAIIRSSDRGATWSFTNLPKVGGMNP	143	
QY	121	GRGAGERLAVDPANSNITTYGARGSGNLKSTDCGVTFSVSSFTATGYTIPDPSNGY	180	
DB	144	GRGGERLAVDPANSNITTYGARGSGNLKSTDCGVTFSVSSFTATGYTIPDPSNGY	201	
QY	181	NSDKGIMVWTFPSTSTGATSRIFVGPADNTTASVYVSTAGSIVSAVPOPG	KYF 239	
DB	202	TSDPVGIAMWTFPSTSTGATSRIFVGPADNTTASVYVSTAGSIVSAVPOPG	KYF 260	
QY	240	PHRAKIDPAKALVLYSDGTGPDGLGSVMRYDINGATWKDITTPVSGSDLYGFEGLG	299	
DB	261	PHKGVLSPEKTLTISVANAGPYDGTNGVHKYNTISGVWTDSPISLASTYVYGLS	320	
QY	300	LDLQPGTLLVAVSINSKMPDQFLRSTDSGTWSPITWAMASPYETIYYISIPKAPWK	359	
DB	321	VDLQPGTLLVAVSINSKMPDQFLRSTDSGTWSPITWAMASPYETIYYISIPKAPWK	380	
QY	360	NNFDVYSESPDGLIKRLQWMIETLEIDPDSNHLVXGGMITFGGHDLTNNDTRNVS	419	
DB	381	---DTTS---TDOFPVGVGMVEALAIIDPDSNHLVXGGMITFGGHDLTNNDTRNVS	433	

```

OY 420 IOSLADGIEEFSYODILASAPGSELLAAVDDNNGFTFASRNDGTSPOUYMAPTWATST 479
Db 434 VKSLAVGIEEHAVALGLITIPGPRALLSAVDDGCF---YHSDLAAPNOAHHTPIYGTIN 490
OY 480 SVDIAGNSVKSVRVNGTAC-TOVAISSDGCATWISIDYAADTSMNGGTVAYSADGDTILM 538
Db 491 GIDVAGNPSNIVISGASDDVPTLALSNCGSTWYADYAASTGTGCAVALSADGPTVL 550
OY 539 STASGVOORQFQSPASVSLPAGAIYASDKTKNSVFYAGSGSTFYVSKDNDSSTTRCP 598
Db 551 MSSTISGALVSKSQGLTRAVSSLPGAVIASDKSDNVTYFGSGAIIYVSKNPAITSTFKV 610
OY 599 KLGSAGITRDLAAHPPTAGTLYVSTDGIEIRSTDSGTFEGOVSTALTNTYOTALVAGS-- 656
Db 611 SLGSSSTYVNAIRAHPSIAGOVMASTDGKLHSHDYDSTFTQIISGVTACHSGFQKASST 670
OY 657 GSNMNLVAFGTGPGSGARLYASDGSASWTDI-QGSGQFSGIDSTKYAGSGSTNAGOVYCT 715
Db 671 GSYVYIYGFPTIDAGLFEKSEDAGTMMOYISDASHQFSGSANVYNGDLOTYGRVFRSH 730
OY 716 NGRGFGFYAQ-----GTVGGTGTGT-----SSSTKSSSSSTSSASSFTLR---SS 757
Db 731 ERPGHLLRQSRREPARGHRGDDGDTTTSKTSSTVSTTLTKTTTSASATTSSTSTTVKTTTSS 790
OY 758 VSTTRASTVTSSTRTSSNACPTGSG--VAGHYAACGGIGIAGMTGPTQCVAPYVQOKONDY 814
Db 791 SSTTSKASSTTTTKTTTSTTSSGTTATASAAQCGGNCMTGATVCFTGTCYENAEY 850
OY 815 YOCV 818
Db 851 SOCV 854

```

RESULT 2

Probable secreted cellulase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: J35227
 R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrel, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: 221572
 A:Accession: J35237
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-890 <SEE>
 A:Cross-references: EMBL:AL031515; PIDN:CAA20642.1; GSPDB:GN00070; SCOEDB:SC5C7.300
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SC5C7.300

Query Match	36.6%	Score 1591.5	DB 2	Length 890
Best Local Similarity	39.9%	Pred. No. 8.6e-77		
Matches 345	Conservative 134	Mismatches 286	Indels 99	Gaps 23

[illegible]

Db 274 TGYLAKGVLAENKYLVLAVSDTGGPYDGGKGRLYRATATGTMTDISPAAEADTYGCF 333

Qy 236 GGLGLDLOKPEGLLVVASTLSNWPDAQLFRSDSGTWSPIYAMASYPIETTYIYSTIPKA 355

Db 334 SGLTLDROKPRGVMTAAVSNWMPDQIFRSDSGATWSQAMSYSTSPRENRKTYMDSSS 393

Qy 356 PMIKNNFIDVYSESPDGLIKRLGMIESLEIDPDSNHHLYLXGCMTFJGGHDLTNDOTR 415

Db 394 PMLTGW----ANPAPPEQTPK-LGMHTALELIDPDSRMMKYGATYGTENELTNMDE 448

Qy 416 HNV-SIGSLADGIEEFVSQDILASAPGSELLAAVDDNGFTFASHNDLGTSPOYVATPT 474

Db 449 GGTFAVEDVWVGLLETTAVNDLASPPSGAPLLSALGDVCGFRHTSLTEV---PSMYTSPN 505

Qy 475 WATSTSVYIAGNSKSVYRVGN-TAGTOVALISSDGGATWSIDYADTISMGCTAYAISDG 533

Db 506 FTTSTSLDFAETKPRDVYVRAAGNLDSGPHIAESTDNGAMFMFGTDPSSGVSGGTAAAGADG 565

Qy 534 DTILWSTASSGVO--RSQFSGFASVSLPACAVIASDKNTSVFAGSGCFYVSKDPS 592

Db 566 SRFVWSPPGAGVQYITGGTSGWASTGLPRACAIYESDKNVNRATYGFPSGFRFYSTDGCA 625

Qy 593 SFTTRPKLG-SAGTIRDIAIHPTTAGTLVYSTDV----GIFRSTDSGTPFGQ---VSTA 643

Db 626 TETASAAATGLPAGDGVREKALPRGEGDVLWLAGAADRGYGLMHSIDGGGFTPLRPGVDAA 685

Qy 644 LNTNYQLALGVGS---GNNMLYLAEFTGPRSGAR-LYASGDSGASMTDQSGSGGSIDST 699

Db 686 DT-----VGGKAPKASVYQTLFTSAELGGVRIEFTSDAGATWTRVNDDAHQMGWTGA 739

Qy 700 KVAGSGSTAGCAGVYVCTNRCGVFYAQ-GTVGGGT-----CG 733

Db 740 ALITGPRRYVGRKYVATNRCGVLYIGDTSDTGGCTDGPGRDPTPGACAEVYTYVYNQMGCG 799

Qy 734 TSSSTKQSSSTSS-----ASSSTTLRS---SVVSTTRA 764

Db 800 FOADYRLNTNTGTSANWNSLDMSPFGQGEVTRMMWNAEHTQAGTSYTAARNVGNMAGVAPGA 859

Qy 765 ST---VTSRTSSAAGPTGSGVAG 785

Db 860 SVRGFTGSRGTNAEPGFAVAG 883

RESULT 3

Probably secreted sialidase, several ASP-boxes and dockerin domain [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: D97013

R:Rolling, J.; Britton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R., J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: D97013

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-839 <R>

A:Cross-references: GB:AE001437; PIDN:AAK78895.1; PID:g15023820; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC0919

Query Match	34.5%	Score 1500.5	DB 2	Length 839
Best Local Similarly	41.7%	Pred. No. 5.2e-72		
Matches 313	Conservative 121	Mismatches 268	Indels 49	Gaps 21

[illegible]

```
Db 96 TMLGESIATPIDITNRVYIAAGLYTNDMDENAYILLSQDKGNTWKRYOLFVKVGNMP 155
Oy 121 GRGAGERLAVDPANSNIITYFGARSGNGLMKSTDGVTFSKSSFTATGTYIPDPSDNGY 180
    || ||||| :|| :|| ||||| ||||| :|| ||||| :|| ||||| :|| ||||| :||
Db 156 GRNMERLQIDPNNDKILYLKARSGNGLMKSEDTGQYTSKVDNFPDITGDIYDP--QNEY 213
Oy 181 NSDKGLMWTFFDSTSTTGATSRIFVGTADNITASVYSTNAGSTWASVAPGQPKYFP 240
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 214 TADKGVYWEFFDSTGKSGSPDTQMYGAADKGNINITYTNDGKTSWASVAGQPKYLP 273
Oy 241 HKAKIQPAEKALYLYTSQGTGTYDGLSVKRYDIAGTMDITVSGSDLYFGGGL 300
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 274 HHGILL-ASDGMILYISYTCGPDGSDQVWKYNTKTEMTNITPAVDITSGSGGISY 332
Oy 301 DLQKRGTLVVASLNSMMPDAOLFSTDSGTTSPIMAMASYPTETYYYSISPKAPWIKN 360
    || ||||| :|| ||||| :|| ||||| :|| ||||| :|| ||||| :|| ||||| :||
Db 333 DAQNNNNVYATLNRMWDEEIRYRSTDAKTKTPIMDMNGYINRLGLNLDYSAQPMIDW 392
Oy 361 NFIDVTSSEPSDGLIKRLGMWIESLEIDPTDSNHLVGTGMTIFGCHDLTNMDTRHNSI 420
    || ||||| :|| ||||| :|| ||||| :|| ||||| :|| ||||| :|| ||||| :||
Db 393 GKTGYT---PPDPLVK-LGMMWGLDEIDPFNSDRMFYTGATLGTDLTNMDKKNVDI 448
Oy 421 OSLADGIEEFYQDLASAPGSELLAAVODNGFTFASNDLGTSPQYVATPTMATSTS 480
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 449 SVKANGIECAVNDVVPYPTKGAQLLSAVGDDCGF---YHDDITKVPKMTTPNESATTS 505
Oy 481 VDYAGNSVSVYRVGN--TAGTQ---VAISSDGAATWSIDYAADTSMNG---GTVAIS 530
    :|| :|| :|| ||||| :|| :|| ||||| :|| ||||| :|| ||||| :|| |||||
Db 506 IDYASVPEFVYRVGNVNDTSKMQQDKCISTDGGKNW---FSAQSNISGVYKAGTVAAG 562
Oy 531 ADGDITLMS--TASSGVORSOFQGSFASVSLPAGAVIASDKRTNSVFAVSGSTFYYSK 588
    || ||||| :|| :|| ||||| :|| ||||| :|| ||||| :|| ||||| :|| |||||
Db 563 ADAKTIVSPPEGANAAYSTDNKMKTPCQSLPGQAKVRSMDVNRKFFGLNGKFYIST 622
Oy 589 DTGSSFTKPKLG----SAGTIRDLAHPHTAGTLYVSTDV-GIRSTDSGTTFCGVSTA 643
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 623 DAGAFTTQSSQGTGKPTKKGIRFTYIGH--EGDIWIAAGKGLMHSSTDGATFFKVSQV 679
Oy 644 LNTTQIALGVSGSGSMN---LVAFGTGPGSARLYASGDSGASWTDI-QGSGQGSIDS 698
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 680 DASD---TVGLQKSKTDDCYPAIYMDATIDGTAGLIFRSDDEATWVRINDAHQIGSPDY 736
Oy 699 TKVAGSGTAGOVYVGTNGRGVFAVAGTYVG 729
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 737 C-ITGDPNKYGRVFGVGTNGRGIVY--GDIDG 764
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 4
F72393
Hypothetical protein TM0305 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: F72393
R:Neelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A12200; MUID:99287316; PMID:10360571
A:Accession: F72393
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-707 <ARN>
A:Cross-references: GB:AE001712; GB:AE000512; NID:94980799; PIDN:AA05393.1; PID:9498080
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0305

Query Match 20.7%; Score 899; DB 2; Length 707;
Best Local Similarity 32.7%; Pred. No. 2.8e-40;
Matches 248; Conservative 118; Mismatches 282; Indels 110; Gaps 33;

Oy 2 FSWKNVNLGGGGFVPGIIFHPKTRGAVAYARRDIGLYRLNADDS-WTAVTDCI-ADNAG 59
```

```
Db 21 FEWKSVEI-NGGCFVPGIIFHPASPLLYARTDVGILRWDEETKRMKOLFDFLRDQSD 79
Oy 60 WHNMIDVAILDPODDQKYAAVAMYTNSMPSNCAITRSSRGATWSTNLP----FKVG 116
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 80 Y--MGVLSAALDPSDPKRYAATGKRYQDM-AGYALILISEYGETWIVLADKXGIVG 136
Oy 117 GMPRGAERLAVDPANSNIITYFGARSGNGLMKSTDGVTFSKSSFTATGTYIPDPSD 176
    || ||||| :|| :|| ||||| ||||| :|| ||||| :|| ||||| :|| ||||| :||
Db 137 GNEEDRNAGERLQVDPNNSVLPKGT-TKYLKMKSEDEGKMKKKYDSTPST----- 186
Oy 177 SNGVNSDKGLMWTFFDSTSTTGATSRIFVGTADNITASVYSTNAGSTWASVAPGQPG 236
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 187 -----SVTFVLPDEKSEKSGPTPRIFVGCSE--PKGIFVTEDEGCTWNVLPNLPN 235
Oy 237 KYFPKAKIQPAEKALYLYTSQGTGTYDGLSVKRYDIAGTMDITVSGSDLYFGGGL 296
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 236 DLPLRGKIH--DGLIYVTLASNALGPNCATRGAVKRYIADQKWYDVPMKGD-----FG 288
Oy 297 GLGLDLQKRGTLVVASLNSMMPDAOLFSTDSGTTSPIMAMASYPTETYYYSISPKAP 356
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 289 YGIDVQE-NVYIVSTLDRWYPHDEIFISLNGETWRPLEKANPDIN-----KAP 338
Oy 357 WIKNNEIDVTSSEPSDGLIKRLGMWIESLEIDPTDSNHLVGTGMTIFGCHDLTNMDTRH 416
    ||||| :|| :|| ||||| :|| ||||| :|| ||||| :|| ||||| :|| |||||
Db 339 WIK-----DLNPH-----W-ISDYKIDPFDMNRAIFTTGCVWVYELK----- 377
Oy 417 NVSISLAD-----GIEFYSYQDLASAPGSELLAAVODNGFTFASNDLGTSPQ 467
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 378 --SFGCMKPKVYKIFENRLEETVVLQVLPPIGEPRLLSALADMGGFRHES---LDTPRS 432
Oy 468 TVMATPTMATSTSYDAGNSVSVYRVGNTAGTQVAISDGGATV-SIDYAADTSMNG- 525
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 433 SMYKPLK--TSLGIAFAIQNSKFVARVHTYTPELSYSDGGINMKREIETVEGTLTDGR 491
Oy 526 -TVAYSADGDTLMTASSGVORSQFQ--SFASVSLA--GAVIASDKRTNSVFA-- 578
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 492 ISLAVNSDQKTLWSPANHEVIVSSDKSKMKKALISVPEFENYPPASDPVPSKIFYED 551
Oy 579 GSGSTFYYSKDTGSSFTKPKLG----AGTIRDLAHPHTAGTLYVSTDV-VGIFRSD 632
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 552 WKNGDFLISKDGKSFMKAKLPSPDNMWSLYSPVLAAPDREGIWLALQNMGLYRSKD 611
Oy 633 SGTTF---GVSTALTNVYQIALGVGS-GSNW-NLYAFGTGPGSARLYASGDSGASWTDI 687
    ||||| :|| :|| ||||| :|| ||||| :|| ||||| :|| ||||| :|| |||||
Db 612 GGITFERLGNVDA----YVIGFAGPKPGTQDYPALIYNKMNAGVYGIPTSDGKTMKRI 667
Oy 688 QGSG--GFGSIDSTKYVAGSGSTAGOVYVGTNGRGVFAQ 724
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 668 NNDKHQFGWIH--YMGDMNEFGRIFLGTGEGRGIVGE 703
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 5
T34434
Hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisels, C.; Gatlung, S.
A:Submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: 221525
A:Accession: T34434
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GEI>
A:Cross-references: EMBL:U08046; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 20
```


sericin1B - silkworm
C:Species: Bombyx mori (silkworm)
C:Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 09-Sep-1997
C:Accession: S52714

R:Garel, A.A.; Deleage, G.G.; Prudhomme, J.J.
Submitted to the EMBL Data Library, March 1995

A:Description: Structure and organisation of the Bombyx mori sericin I gene and of the

A:Reference number: S52714

A:Accession: S52714

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1217 <GAR>

A:Cross-references: EMBL:248802; NID:q755699; PID:q755700

Query Match 6.1% Score 267; DB 2; Length 1217;
Best Local Similarity 22.2%: Pred. No. 1.5e-06;

Matches 175; Conservative 106; Mismatches 306; Indels 202; Gaps 31;

```
QY 31 ARTDIOGLRLNADDSMTAVT-DGIADNAGMHWGIDAVALPDODQKYAAVGMVTSW 89
DB 544 AGRAGGATKSEASSSTQATTVSGADDSADSYTW-----W 579
QY 90 DSNCAIITSSDRGATWSTNLPFKYGMPCRGAGERLAVDPANNTIIFGARSGNGLM 149
DB 580 NPR-----RSSSSSSSSSSSSSGSNVGS--SQSSGOSTSGSNARGHL---GVSSSTGST 629
QY 150 KPDGCVTFKVSFATGTIIPDPSDNGYNSDKGLMMVTFDSTSTGATSRIFVG 209
DB 630 SMTD-----SSKSACSSRTSGGTSTYGYSSSHRG-----GSVSSV-----G 665
QY 210 TADNITASYVSTNAGSTWSAVPGOGPKYFPKAKLQPAEKALYLYSDGTGPDGLGS 269
DB 666 SSSNTDS--TKNAGSSTSGGTSTYGYSSSHRG-----GS 698
QY 270 VWRVYDIAGTWADITPVSSSDLYFGGGLDLQKPTLVASLNSWMPDAOLFRTDSC 329
DB 699 VSSSTGSSNT-DSTKSAAGSSTSGGTSTYGYSSRHARGAVSTGSS-----SNTDSS 749
QY 330 TWPSPITMANASYPTERYYSISTPKAPWIKNFIIVTSESPDGLIKRLGMWIESLEIDP 389
DB 750 T--KNAGSRTSGGTSTYGYSSSH--RGGSVSTGSSN----- 783
QY 390 TDSNHWLYGTGMITFGHDLTWMTRHNYSIOSLADIEEFVODLASAPGSELLAAGV 449
DB 784 TDS--TKNAGSRTSGGTSTYGYSSSH-----GGS--VSSTG 817
QY 450 DONGFTFARNDLGTSPTQVMAPTWATSTSYDVAGNSYKSVYRGVNT-AGTOVAISSDG 508
DB 818 SSSNTDSSSTKN--AGSRTSGGTSTYGYSS--HRGGSVSTGSSSNTDSSSTKNAGSSTS 872
QY 509 GATWISIDVYAAD-----TSMNGGTVAYSADGDTILMSTASSGV 545
DB 873 GGSSTYGYSSDSRDGVSSTGSSSNTDASTDLAGSSTSGGSSTYGYSSDSRDGVSSTGS 932
QY 546 ORSOFQGSFASVSLPAGAV---IASDKTNSFYFAGSGSTFYVSKD-TGSFTRGPK-L 600
DB 933 SSNTDASTDLAGSSTGGSSTYGYSSDSRDGVSSTGSSSNTDASTDLGSSSTGGSSTY 992
QY 601 GSAGTIRDLAAPT-TAGTLVYSTDVYGFIRSTDSGTTFCQVSTALTNTQIALGVGSGN 659
DB 993 GYSYSDRDSVSSTGSSSTDLASTDLAGSSTGGSSTYGYSS--SNRGSVATSSSSN 1049
QY 660 W-----NLVAFGTGPGSARLYASGDSASWMDIOGSGFGSIDTKVAGSGSTAGQYV 712
DB 1050 TDASTTEESTTSGSSTEG---YSSSHDGSVYSTGSSSTSGCASSSSASTASDA----- 1102
QY 713 YSTNGRVFYAOGTVGGTGTSSSTKQSSSTSSASSSTTLNLSVYSTTRASTVTSST 772
DB 1103 -ASSEGGFWMNRRKSG--SGHKSATVQSSSTDKTSTDASSSTDSSTSSGASTTSSGSS 1159
QY 773 SSAGPTGS 781
DB 1160 STSGSSTS 1168
```

RESULT 8

JU0150

cellulose 1,4-beta-cellulobiosidase (EC 3.2.1.91) I - Penicillium janthinellum

N:Alternate names: exo-cellulobiohydrolase

C:Species: Penicillium janthinellum

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: JU0150; S21508

R:Christoph, A.K.; Weigel, T.O.; Schultz, G.

Gene 124, 57-65, 1993

A:Title: Cloning, sequencing, and heterologous expression of a cellulase-encoding cDN

A:Reference number: JU0150; MUID:93178976; PMID:840481

A:Accession: JU0150

A:Molecule type: mRNA

A:Residues: 1-537 <CHR>

A:Cross-references: EMBL:X59054; NID:q3177; PID:CAA1780.1; PID:q3178

A>Note: the authors translated the codon ATG for residue 172 as Asn

C:Genetics:

A:Gene: cbh1

C:Superfamily: cellulose 1,4-beta-cellulobiosidase I; fungal cellulose-binding domain h

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:506-537/Domain: Fungal cellulose-binding domain homology <FCB>

Query Match 6.1% Score 264.5; DB 1; Length 537;
Best Local Similarity 22.6%: Pred. No. 7.3e-07;

Matches 145; Conservative 68; Mismatches 226; Indels 203; Gaps 27;

```
QY 243 AKLQPAEKALYLYSDGT-GPYDGLGSVWRVYDIAGTWMKIDITPVSGS-DLYFGGGLGL 300
DB 33 AETHRA-----LTWSKCTAGXCSQVGSV-----VIDANPXYHSTGSGTNCYTG----- 77
QY 301 DLQKPTLVASLNSW--WPD-----AQLFRSDGCTTWPSPITMANASYP 342
DB 78 -----NWDATATLCPDVDTCAANCAVDGARROHLRTVTSGN----- 113
QY 343 TETYYYSISTPKAPWIKNFIIVTSESPDGLIKRLGMWIESLEIDPTDSNHWLYGTGWT 402
DB 114 -----LRINFTVTTASQ-----KNIGSLYILEND-----T 138
QY 403 IFGCHDLTWMTRHNYSIOSLADIEEFVODLASAPGSELLAAGVDDNGFTFASRNDL 462
DB 139 TYOKFNLNDEFFEDVDVSNLPCGL-----NGALFYVDMDADGMAKYPPTNKA 186
QY 463 GSPQVMAPTWATSTSYDVAGNSYKSVYRGVNTAG-----PQVATSSDGGATWISD 515
DB 187 GAK-----YGTGYCDSCQCPRLKFTINGQANVDGWTSPKNDVNSGIGNHSCCAEMD 237
QY 516 YIADTSMNGGTVAYSAD--GDTILMSTASSGVORSQFQGSFASVSLPAGAVIASDKTN 573
DB 238 IMEANSISNAVTPHPCDTPSQMCTGQRCGCTYSTIRYGTCD---PQGCDFNPYRMGV 293
QY 574 SVFYAGSGSTFYVSKDGTSGSFTRGPKL-----GSAGTIRDLAAPTAGTLVYSTDVGIT 628
DB 294 TNEY-GPGETI---DTKSPFTVYQFLTNDGTSTLTSEIKRFYVQGGKV-----IGNP 343
QY 629 RSDSDGTTCQVSTALTNTQIAL-----GYGSG-----SNWN----- 661
DB 344 QSTTVGVNSGNTDSCNOKSAFGDTNEPSKHGAGAGAGLADGMLVMSLMDHASD 403
QY 662 -LYAFGTGPGSARLYASGDSASWMDIOGSGFGSIDST---KVAGSGSTAGQVYVGTN 716
DB 404 MLMIDSTYPTNA---ISTTPGAKRGTCDSLRNRYESTYPPNAYLYVSNIKTKGPL----- 455
QY 717 GRGVFYAOGTVGGTGTSSSTKQSSSTSSASSSTTLNLSVYSTTRASTVTSSTSSAA 776
DB 456 -----NSTFTFGGTSSSSTTTTTSKSTSSSSKT-----TTVTTTTSSTSSG 498
QY 777 GPTGSGVAGHYAOCGIGMTGPTQCAVAPVCOKONDYVYOCV 818
DB 499 SGTG---ARDWACQGGNGWTGPTTCVPTTCKONDMYSOCL 537
```

```

RESULT 9
mucin, submaxillary - pig
N:Alternate names: apomucin
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 02-Jun-2000
C:Accession: T03099; A40009; A28528; B29789
R:Eckhardt, A.E.; Timppe, C.S.; Deluca, A.W.; Hill, R.L.
J: Biol. Chem. 272, 33204-33210, 1997
A:Title: The complete cDNA sequence and structural polymorphism of the polypeptide chain
A:Reference number: 214839; MUID:98070526; PMID:9407109
A:Accession: T03099
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1328 <ECC>
A:Cross-references: EMBL:AF005273; NID:92581863; PIDN:AAC62527.1; PID:92581864
R:Eckhardt, A.E.; Timppe, C.S.; Abernethy, J.L.; Zhao, Y.; Hill, R.L.
J: Biol. Chem. 266, 9678-9686, 1991
A:Title: Porcine submaxillary mucin contains a cysteine-rich, carboxyl-terminal domain in
A:Reference number: A40009; MUID:91236743; PMID:2033060
A:Accession: A40009
A:Molecule type: mRNA
A:Residues: 12139-12167, 'T', 12169-13288 <ECC>
A:Cross-references: GB:M61883; NID:9454837; PIDN:AAA30998.1; PID:9164374
R:Timppe, C.S.; Eckhardt, A.E.; Abernethy, J.L.; Hill, R.L.
J: Biol. Chem. 263, 1081-1088, 1988
A:Title: Porcine submaxillary gland apomucin contains tandemly repeated, identical sequ
A:Reference number: A28528; MUID:88087170; PMID:2826455
A:Accession: A28528
A:Molecule type: mRNA
A:Residues: 12139-12167, 'T', 12169-12641 <TIM>
A:Cross-references: GB:M21174; GB:J03512; NID:9164321; PIDN:AAA30990.1; PID:9552360
R:Experimental source: submaxillary gland
R:Eckhardt, A.E.; Timppe, C.S.; Abernethy, J.L.; Tonnadje, A.; Johnson Jr., W.C.; Hill,
J: Biol. Chem. 262, 11339-11344, 1987
A:Title: Structural properties of porcine submaxillary gland apomucin.
A:Reference number: A92606; MUID:87280230; PMID:3611111
A:Accession: B29789
A:Molecule type: protein
A:Residues: 1572-1607 <EC2>
C:Superfamily: pig submaxillary mucin
C:Keywords: tandem repeat

Query Match 6.1%; Score 263; DB 2; Length 13288;
Best Local Similarity 24.3%; Pred. No. 4.7e-05;
Matches 213; Conservative 97; Mismatches 399; Indels 166; Gaps 35;

OY 11 GCGGFPGLIIFHKTKGVAARTDGLRLNDDSWTAVTDCIADNAGHNMGIDAVAL 70
DB 1636 GSPGATGASIGOPETSRISVAGS--SGAPAVSSGASQAAGTSGAGPQTASSVGTETAR 1693
OY 71 DPDDQKYVAAGMYTNSMDPSNG-----ATIRSSDRGATWSFTNLRF 113
DB 1694 PSVAGSGTTGTVSGAGSGTSSGSPGATGASTIGOPETSRISVAGSSGAPAVSSGASQA 1753
OY 114 KVGGNPGRGAGERLAVDPANSNIITFGARSGGLMKSTDC--GVTFKVSFTATGYI 171
DB 1754 GTSGAPGPTTASSVGTETARPSV-----AGSGTTGTVSGAGSGTSSGSPGATGASI 1807
OY 172 PPSDSNGYNSDKQGLMWTFTDST-----SSTTGATSRIFVGTADNITASY-----YV 220
DB 1808 GOPETSRISVAGSSGAPAVSSGASQAAGTSGAGPQTASSVGTETARPSVAGSGTTGTIV 1867
OY 221 STVAGSTWSA--VPGQPGKYF--PHRAKLOPAEKALYLYTSDGTGYDGLGS----- 269
DB 1868 SGAGSGTSSGSPGATGASIGOPETSRISVAGSSGAPAVSSGASQAAGTSGAGPQTAS 1927
OY 270 -----VMRYDIAG--GTWKDITPVSSDLYFGF-----GGLGLDLOKPGTLVVA----- 311
DB 1928 SVGTETARPSVAGSGTSSGASGSGT--GSSSGSPGATGASIGOPETSRISVAGSSG 1984
OY 312 --SLNWMHPDAOIFRSTDSGTTSPITMANASYTETLYTYSISIPKAPWKINFDVTSBS 369

```

```

DB 1985 APAVSSGASQAAGTSGAGPQTASSVGV-----TETARPSV-----AGS 2023
OY 370 PSDGLIKRLGMMIESLEIDPTDSNMHLVGTGMITFGCHDLTMDFRHNYSIOSIADGICE 429
DB 2024 GTTGTVSGAGSGTSSGSGSG-----ATGAST--GQPET--SRISVAGSSGAPAVSS 2071
OY 430 FSVQDASAPGSGSELLAAVDDNGFTFASRNDLGTSPPQVMAITPTWATSTSVYDAGNSVK 489
DB 2072 GASQAAGTSGAGPQTAS--SVGTETARPSVAGSGTT--GTVSGASGSGTSSGSPGA 2126
OY 490 SVYRVNCTACTQVAISDDGA---TWISIDYADTSNMG--GYAVASND--GDTILMTASSG 544
DB 2127 TGASIGOPETSRISVAGSSGAPAVSSGASQAAGTSGAGPQTASSVGTETARPSVAGSG 2186
OY 545 VQR--SQFGSPFASVSSLPAGAVIAS--DKNTSVYAGSGSTFFVYKDTGSGSFTGPRL 600
DB 2187 TTGTVSGAGSGTSSGSP--GATGASIGOPETSRISVAGSSGAPAVSS--SGASQAAGTSG 2243
OY 601 GSAGTIRDI-----AAPTTACTLVYSTVDVGFIRSTDSGTTGQVSTALTNVQIALVYG 655
DB 2244 AGPQTASSVGTETARPSVAG-----SGTT--GYVSGAGSGT-----GSS 2282
OY 656 SGGNMNLVAFGTCPSGARLYASGDSGASMTDIOGSGFGSIDTKVAGSGSTAGQVYVGT 715
DB 2283 SGSPGATGASIGOPETSRISVAGSSGAPAVSSGASQAAG---TSGAGGPTTASSVGYRE 2338
OY 716 NGRGVYACQTVG---GGTGGTSSSTKQSSSTSSASSTLLRSSVYSTRASTVTSRT 772
DB 2339 TARPVAGSGTCTTVSGAGSGTSSGSPGATGASTIGOPETSRISVAGSSGAPAV--SSGA 2397
OY 773 SSAAGPTGSGVAGHYAQC-----GIGWGT 797
DB 2398 SQAAGTSGAGPQTASSVGTETARPSVAGSGTTG 2432

RESULT 10
S11672
Ice nucleation protein - Xanthomonas campestris
C:Species: Xanthomonas campestris
C>Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999
C:Accession: S11672
R:Zhao, J.; Orser, C.S.
Mol. Gen. Genet. 223, 163-166, 1990
A:Title: Conserved repetition in the ice nucleation gene Ixh from Xanthomonas campe
A:Reference number: S11672; MUID:91080859; PMID:2259339
A:Accession: S11672
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1567 <ZHA>
A:Cross-references: EMBL:X52970; NID:948531; PIDN:CAA37140.1; PID:948532
C:Superfamily: Ice nucleation protein

Query Match 5.9%; Score 257.5; DB 2; Length 1567;
Best Local Similarity 24.2%; Pred. No. 6.5e-06;
Matches 223; Conservative 87; Mismatches 400; Indels 213; Gaps 42;

OY 24 KTKGVAYARTDIGGLYRLNADDSMTAVTDGIADNAGHNM-----MGIDAVALLDPDDQKY 78
DB 200 KTFETGYGSLTG-----ADGSRVLAGVGTETACDHDLLAGYSGTGA---GSDSSI 250
OY 79 YAAVGMVYNSMDPSNCAIIRSSDRGATWSFTNLPEKVGNNMPGRGAGERLAVDPANSNII 138
DB 251 LAGYG-----STQTAAGRSTLTAGYSGTQTAOE-----CSRLLSGYGSATSGSDSAVIS 300
OY 139 YFGARSGNGLMKSTDCV-----TFKVSFTA---TGTYIPDPDSNNGYNSDKQGLMNV 190
DB 301 GYGSTQTPAGESSELTAGYSTQTAARKGSDITAGYSGTGTAGSDSALILAGYSTQ----- 354
OY 191 TFDSTSTGATSRIFVGTADNITASYVSTNA-----GSTSAVPGQPGKYF 239
DB 355 TAGSESSLTAGYSGTGTAKAGSDVYTAG--YGSTGTAGADSTLLAGYSGTGA---GGSSSLT 411
OY 240 PHRAKLOPAEKALYLYT---SDGTGYDGLT---GSVWR---YDIAGTGWKIDITPVSG 288

```

Db 412 AGYGTOTAROGSDITAGYGTGTAGADSTLLIAGYGTGTGSDSLTAGYGTOTARKG 471
OY 289 SDLYGFGGLG-----DLOKPGTLVASLNSMMPDADLPSTDSGCTW---- 332
Db 412 SDITAGYGTGTAGSDSLIAGYGTGTAGS--ESSLTAGYGTOTAROGSDSLTGGYST 529
OY 333 -----SPIWMAASYPTETV-----YYSISTPKAPWIKNFIIDVISESPSDG-LIKR 377
Db 530 STAGHDSLLIAGYGTGTAGYDSTLLIAGYGTGTAGDSTLLIAGYGTGTAGDSTLLIAG 589
OY 378 LGMWLESLEIDPTDSNHL-YGTGWTIFGHDLT-----NMDTRHNVSIQSLADGLEEFSV 432
Db 590 YG-----STGTAGSDSLTAGYGTGTAREGSDVYTAGYGTGTAGDSTLLIAGYGTGTG 645
OY 433 ODLASAPG-GSELLAAVDDNGFTFASRNDLGTSPQTV---WATPMATSTSVDAAGNSV 488
Db 646 SDSSLTAGYGTGTARKGSDVYTAGYGTGTAGADSTLLIAGYGTGTGSDSLT-AGYGS 704
OY 489 KSVRVGN-----TACTOVAISSDGCATWSIDYAADTSMNG----- 525
Db 705 TOTARKGSDVYTAGYGTGTAGADSTLLIAGYGTGT-SSDSSLTAGYGTGTARKGSDV 762
OY 536 TVASADCDTILMSTASSGVORQFQSGFASY-----SLPAGAVIASDKKT 572
Db 763 TAGYGTGTAGADSTLLIAGYGTGTGSDSLTAGYGTGTARKGSDITAGYGTGTAGA 822
OY 573 NSVFAGSGSTFYVSKDQSSFTRGPKIGSACTIR--DIANHTACTL-----Y 620
Db 823 DSTLIAGYGTGTGSD--SLTLAG--YGTGTAREGSDVYTAGYGTGTAGADSTLLIAGY 878
OY 621 VSTDVGFIRSTDSGTT--FGOVSTALTNVYQIALGVSGSNMNLAFGTGPGSA--RLY 675
Db 879 GSTGTAA--GSDSSLTAGYGTGTARKGS-DVTAGYGS-----TGTAGADSTLLIA 924
OY 676 ASGDSGAMTIDIGSGGCGSIDSTKVAGSGSTAGQVYVGTNGRGVYVAGVGGTGTGTS 735
Db 925 GYGSTGTGSDSLTAGYGTGTGTAR-KGSDMTAGYGTGTAG-----ADSTLIAGYGTG 978
OY 726 SSTKQS-----SSTSSASSSTTLRSVSVSTRA---SVVTSR 771
Db 979 TSGSDSLTAGYGTGTAREGSDVYTAGYGTGTAGADSTLLIAGYGTGTGSDSLTAGY 1038
OY 772 TSSAAGPTGSGVAGHYAOCGGIG 794
Db 1039 GSTGTAROGSDVYTAGYGTGTAG 1061

RESULT 11

F90696
hypothetical protein Ecs0542 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90696
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5291 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA833965.1; PID:913360000; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: Ecs0542

Query Match 5.9%; Score 257; DB 2; Length 5291;
Best Local Similarity 20.8%; Pred. No. 3.1e-05;
Matches 223; Conservative 133; Mismatches 345; Indels 370; Gaps 54;

OY 10 GGGGFTVPGIIFHPRKGVAVARTDIGLYRLNADDSWT-----AVTDGIA----- 55

Db 3562 GTGTNFACTVTVLLNKGYSAT-----IQSGMSVNVNPAVDVALSDGTSTVSA 3614

OY 56 ---DNAGHNNCIDVVALD-----PODDQ----- 76

Db 3615 SAODSAGNSATASRSVADLAPVISINTVSTDDRLNAAE000PLTLNGSTSAEYGVTV 3674

OY 77 -----KVYAAVGMVYNSMPSNGAI-IRSSDGCATWSFTNLPFKYGNMPPRGAGERLAV 130

Db 3675 VTFGKTYTATVAAAGTAAVLANPAYDLAALOGAGCTTASVNDRGN--PGQ-ATHALTY 3731

OY 131 DP-----ANSNIIYFARGSGNLKSTDGVTESKVSSTFATGTYIPDPDSNGY 180

Db 3732 DVAPVTIATVAGDII-----NNAEQLAGCTISGTTAEVGO-TVVTVF----- 3776

OY 161 NSDKGGLMVTDTSSITTGATG-----RIFVGRAD-NITASVYVSTNAGTWSA----- 230

Db 3777 ---NGQTV---SATVSGGMSVFIPAQOPAGLSDSYTSATVSDQACNPGSASRGV 3828

OY 231 -----VPCOPGKYPFHKKALQPAEKALYTVS-DGTGPDGTL-----GSVVRDYI-AGG 278

Db 3829 TLNGDVPYVTITVTFAGDDVYNAEHGSSLYISGTTTAPVGTLLTLNKTITTVTGTG 3888

OY 279 TWK-----DITPVSGDLYF-----GFGGLG-----LDLOKP----- 305

Db 3889 SMSYTLGSADVTALADGNAYVINASVNAIGNTGSSNHTITVDLAPAMGINIDSLQADT 3948

OY 306 -----GLVYASLNSMMPDQFLRSDSGTNSPI-----WAMASFP 342

Db 3949 GLSADFTITSVPVYVNSLTALALASN--ETAOI--SIDGTTWTLTVYGTTPMYNDSR 4004

OY 343 TET-----YYSISTPKAPW-----IKNFIIDVTSSEPS-----DGLIKR----- 377

Db 4005 TLTIDNLYLQVAVIDAAGVAGATDSQNVVIDTADPADPKTTAISAITDKLITINDPVT 4064

OY 378 -----LGMWLESLEIDPTDSNHLVGTGWTIFG-----GHDLTNMDRHNVSI 420

Db 4065 SDPTLAVSGTLGATLAGEFAQISLDGVTWTLTVGVGSMVYADGHTLTDGTWNYTVAV 4124

OY 421 QSLADGLEEFVSQDL---ASAPGSELLAANG--DDNGTFASRNDLGTSPQTVATPIW 475

Db 4125 VDLANVGQTATQANVVDVTTSEPAKSTITGTISDDTG--TSSDFTISDPTL----- 4175

OY 476 ATSTVDYAGNSKSVVRGNAGTQOVAISSDGCATW-SIDYAADTSMNGGVAAVADGD 534

Db 4176 -----TVRGVLAALCANEFADISTDNGCATWVAVTAAD-SLNMSTV-----DGR 4219

OY 535 TILMSTASSGVORQFQGSFASVSL-----PAGAVIASDKKTNSVYFAGSGSTEVY 586

Db 4220 TLTNGTTMQVAVVDLAGVAGTSSQSALIDTVNPAQVLTIASISTDT---GSSATDFI 4275

OY 587 SKDGTSSFTRGKLG-SAGTIDIAAHPTTACTLVSTVDVGFIRSTDSGTTFGQVSTALT 645

Db 4276 TSDIMTLTLLSGLAGLAGEVAOI-----SLDSGATW-----TTLT 4311

OY 646 NTQYALGVSGSNMNLVAFGTGPGSARLY-----ASGDSG----- 681

Db 4312 T-----NOTQWTVYDSRLTLTGSIYQVAVLDLAGNCTPPVSKTVVDFTINPATP 4362

OY 662 --ASWTDIQSGGFGSIDSTK-----VAGSGTAGQ-VYGTNG----- 717

Db 4363 TIVSYTDDVG--QROGTLSSQATDPTPLNLVLAAPLASGEVYLYLRGLLGAVTWVG 4421

OY 718 --RGVYVAGVGGTGTGSSSTKSSSTSSASSSTTLRSVSTTASVTSSTRTSSA 775

Db 4422 ALNMTYSDSGLVSGAVTYTSARVVDLAGNTSSDVLTVVDVTSIPTTLAQIT---SQTTRD 4478

OY 776 AGPTGSGV-----AGHYAOCGGIGMTGPTQCVAP-----YVCOKNDDYVQ 816

Db 4479 TTPILISGVITALASGQYVE---VINKKTYISBEAGAVVNDPAINTWYVQ 4526

RESULT 12

B85547
Probable RTX family exoprotein [imported] - Escherichia coli (strain 0157:H7, substrain C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85547
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lm, A.; Dimalante, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; MID:g1513368; PIDN:AG054838.1; GSPDB:GN00145; UMGCP:206
A:Experimental source: strain 0157:H7, substrain EDL933
C:Genetics:
A:Gene: 20615

Query Match 5.9%; Score 255; DB 2; Length 5188;
Best Local Similarity 20.9%; Pred. No. 3.9e-05;

Matches 224; Conservative 131; Mismatches 347; Indels 368; Gaps 55;

```
OY 10 GGGGCVPEIIFHPRKGVAVARTDGLGLRLNADSWT-----AVTDCIA----- 55
DB 3459 GGTGTEATGTVTYVLLNGKYSAT-----IQSNGSVMVXADVAALSDGTSTYVSA 3511
OY 56 ---DNAGMNMGIDAVALD-----PODDQ----- 76
DB 3512 SNODSGNATSRSAVAVDLTAPVIVISNTDDBKXNAAEQOPLTLNKGSTSEVQYVT 3571
OY 77 ----KVVAAGMNTNMPDSNGAI-IRSSDRGATWSFTNLPFGKGNMPGRCAGERLAV 130
DB 3572 VTFGCKTATATVAANCTMALNPVAVDLALGCGAQTITASVNDRAQN--PGQ-ATHALTV 3628
OY 131 DP-----ANSIIYFGARSGNGLMKSTDGCTFEKVSSTFATGTYIYIDPPSSNGY 180
DB 3629 DTVAPVTITATVAGDIT-----NNAEOLAGQITSGTTAEVQO-TVTYTF----- 3673
OY 181 NSDKOGLMVFDFSTSTTGATG-----RIFVGTAD-NTTASVYVSTNAGTWSA----- 230
DB 3674 ----NQGTW-----SATVSGGSGSVFIPRAQFAGLSDSYTTISATVSDAGNPGSASRV 3725
OY 231 ----VPGQPKYFPKAKLQPAEKALYLYTS-DGTGPDGLT-----GSVMRYDI-AGG 278
DB 3726 TLNGDVPVTINTFAGDDVAVNAEHGSSLVISGTTAPVGGTLTLTLNGKTYTTVQTG 3785
OY 279 TKK-----DTPVSGSDLYF-----GFGGIG-----LDLQKP----- 305
DB 3786 SMSYTLGSAVDVTLADGNVAVINASVSNAGTSSNHTTVDLSAPAMGINIDSLQADT 3845
OY 306 ----GTLVVASLNSMMPDQALFRSTDGTTMSPRI-----MAMASYP 342
DB 3846 GLASDFITSVSPVYVNGSLTALAASN--ETAOI-SIDGTTTWTLYTGTTWRKNDNR 3901
OY 343 TET---YYISISTPAAPW-----IKNFIIDVTSSEPS-----DGLIKR----- 377
DB 3902 TLTGNYLYLQVAVIDAAGVAGATDSQNVVIDTAPPAVKTLAISAITTDMGLITNDFVT 3961
OY 378 ----LGMITESLEIDPTDSNHLXGTGNTIFG-----CHDLTNDTRNNSI 420
DB 3962 SPTTLAVSGTLATLATSAGFAOISLDGGVTTWTLTVVGSWSYADHHTLTDGTWNTVAV 4021
OY 421 QSLADGIEEFSVQDL---ASAPGSELLAAVG--DDNGFTFASRLDGTSPOTVATPTW 475
DB 4022 VDLAGVGTATQANQVVDTPSPAKSITITGIXDTC---TSSSDFITSDTTL----- 4072
OY 476 ATSTSVDAVGNVSVRVGNAGTQVLAISSDGGATW-SIDVAADTSMNGGVAAVSADGD 534
DB 4073 ----TVRGVLAALANEFQAQISTDNGATVAVTAAAD-SLNMVSV---DGR 4116
OY 535 TILMSTASSGVGRSFGSFAVSLSL-----PAGAVIADSKTKNSVFIYAGSGSTFIV 586
```

```
DB 4117 TLTNCTTQVQVAVDLACNVGATSSQSALIDTVNPAQVLTIASISTDT---GSSATPEI 4172
OY 587 SKDTGSSFPGRKLG-SAGTIRIDIAHPRTAGTLVYSTDVGIIFRSTDSCTTGGQVSTALT 645
DB 4173 TSDTHLTGLTSLGACIASEVAOI-----SLDSGATW-----TTLT 4208
OY 646 NTYQIALGVGSSNMNLVAFGTGPGCARLY-----ASGDSG----- 681
DB 4209 T-----NCTOWTLYTDSRFLTLTGSSVYQVAVDLACNTGPVSVKTVVDTINPTATP 4259
OY 682 --ASWTDIOGS--QG-FGSDSTRK-----VAGSGTACQ-VUYGTG----- 717
DB 4260 TVVSTYTDVQGRQGLTSLQATDPTTLNLGVLAPLASEVAVYLRNGLLCAVTWGA 4319
OY 718 -RGVFAQGTGCGTGGTSSSTKSSSTSSASSTTLRSVSTASTVTSRSTSSAA 776
DB 4320 LMMWYSDSGLVGAGATYTSARVVDLAGNITSSSDPVLTVDTISPTLQAQT---SQTTROT 4376
OY 777 GPTGSGV-----AGHYAOCGIGMTGPTQCVAP---YVCSOKNDYUQ 816
DB 4377 TPLISCVITAAALASGOYE---VINGKTYTSEPRGCAVVVDRAHNTWYUQ 4423
```

RESULT 13

A83412
hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83412
R:Stover, C.K.; Pham, X.Q.; Ertvin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2468 <STO>
A:Cross-references: GB:AE004613; GB:AE004091; MID:g9947856; PIDN:AA05263.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1874

Query Match 5.8%; Score 252.5; DB 2; Length 2468;
Best Local Similarity 22.0%; Pred. No. 2.1e-05;

Matches 195; Conservative 100; Mismatches 333; Indels 259; Gaps 38;

```
OY 7 VKLGGGGE-VPGITFHPKTKGVAVARTDGLGLRLNADSWTAVTDTGADNAGMHNCGI 65
DB 378 VVVGPGSFEVP--LNPLTLNG-----ETVYVIVTDPAGNSSTPV 415
OY 66 DAVALDPODDQVYVAVGNTNMPDSNGAIIRSSDRGATWSFTNLPFVGNMMPGRGAG 125
DB 416 TAEAPDFDAPQVNA-----SNGSVLSGTAE-----AG 443
OY 126 ERLAVDPANSNIYFGARSGNLMKST-----DGV-----TFSKVSEFTATG 169
DB 444 VTIIVTDGNGNPIGOTSADANGMSFTPGSQLPDGVVAVVADAAGNSPATSTLVDGV 503
OY 170 --YIPDSNGYNSDKOGLMWTDFDSTST--TGATSRIFVGTADNITASVYSTNAG 225
DB 504 APNAPVPEPSNG--SELSG-----TAEPGSSVLTGNGNPIQQTAD-----AN 546
OY 226 STWSAVPGQPKYFPKAKLQPAEKALYLYTSDGTGTPVGTGSGVWRVYDIAGT----- 279
DB 547 GNMSTPTSP---LP-----DGVVAVVARDAAAGNSSPASV 580
OY 280 ----WKDITPVSGSDLYEFGGLGLDLOKPGTLVVASLNSMMPDQALFRSTDGTT 331
DB 581 TVDAVAPATPVDPDSNGTTL-----SGTAEPSGSVTLTGNGNPIQV--TADGSGN 630
OY 332 WSPINAMASYPETYIYYISISTPKAPWIKNFIIDVTSSEPSDGLIKRLGMIESLE----- 386
```

```

Db 631 WT-----FTSTPLPNCVTYVNAATADPSGNASSPASYVDAVAPATPV 673
OY 387 IDPTDSNMHLVYGMTIFG----GHDLTWMDPRHNYSIQSLADIEEFVSQDLASAPGS 442
Db 674 VAPSN-----GTLSTAEAGATVTLTDGNGNPIGOVTAADSGSNMFTPTPLPNCGT 725
OY 443 ELLAAVGDNDGCTFFASRNDL--GTSPOVWAMPPTWATSTS--VDYAGNSVSKSVRVGNTAG 499
Db 726 VYNAATATDASGNTASGSSVTVDSPAATPVINPSNCTTILSGTAEPSSVTLTDGNGNPICG 785
OY 500 TQVAISSDGGATWSIDYADTSMNGTVA-----YSADGDTILWSTASSGVQORSQ 549
Db 786 ---QVTADSGNMS--FFPSTPLADGTYVNAATDPAGNTSGOGSTTVGVAPTPPTVM 840
OY 550 FQGSFASVSLPAGAVIASDKTN---SVFYAGSGSTF-----YKMDT--- 590
Db 841 SNKSSLSGTAEFGSTVTLTDGNGNPFAEYVTAADSGSNMFTPTPLPNCVTYVNAADAG 900
OY 591 ----GSSFTKGLGSACTIRIDIAHPPTAGTLVSTDVG--IFRSTDSGTFEGQVSTAL 644
Db 901 NSSPGASVTVDSQAAPAPV-----NPSNGTTLSTAEAGATVTLTDGNGNPICGVTADG 955
OY 645 TMTYQIALG--VSGSNMNLVAFG--TGPSCARLYASGDSGA-----SMTDIOGSOG 692
Db 956 SCNMSEFTPTPLPNCVTYVNAATADPSGNTSAPASTVDSPAAPVNPSCAETISGTAE 1015
OY 693 FGSIDSTKVASGSTAGVYVGTNGCVF-----YAGQVY-----GGGTGTSSTSRK 739
Db 1016 PCA-YVTLTDGSGNPICGVTADSGSNMFTPTPLADGTYVNAATADPAGNTGGOGSTTV 1074
OY 740 Q-----SSSTSSASSSTTLRNVSTTRASTVTSRTSSAAG 777
Db 1075 DALAPATPVNLSNNGSSLSGTAEFGSTVTLTDGNGNPFAEYVTAADSGC 1121

```

RESULT 14

```

T34433
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34433
R:Geisels, C.; Galtung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: 221525
A:Accession: T34433
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1032 <GEI>
A:Cross-references: EMBL:U08046; PIDN:NACT0889.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2; 99

```

```

Query Match 5.6%; Score 243.5; DB 2; Length 1032;
Best Local Similarity 22.4%; Pred. No. 2.1e-05;
Matches 161; Conservative 106; Mismatches 312; Indels 139; Gaps 26;

```

```

OY 128 LAVDPANSNIIFGARS--GNGLMKSTGAGVTFSSFTATGTIIPPPSDNGVNSD- 183
Db 184 LAMDOGNTOKGDLNTHIGNNADTIDINNVLNTATSPSTAVTSPS-SLGTSSSP 242
OY 184 -----KQGLMWTVPDSTSTGGATSRITVGADNITASVYVSTNAGSTWSAVPGQPKY 238
Db 243 LPSISISTSLPIPTASSSSPSAAS-----TTPVLLSSSTIOSSGCTPPSSVASSPSTV 297
OY 239 FHKKAKLQAEKALYLTYSDDGTGPRYTGLGSWRRYDIAGCTKDTIPVGSGLIRFGGL 298
Db 298 ---GSTSGAASSSYATVS-----TLAGSTSTTTPVPGSSSTIGSSTP 338

```

```

OY 299 GIDLQPGTLVAVASINSWMPDAQLFRSTDSGTTWSPINAMASYPTEYYVSTSPKAPW 358
Db 339 SASSSSSSGTM-----STISGSTGTV---TVVPGSSSTFASSTPIA---- 376
OY 359 KKNFDVTSSEPSDGLIKRLGMKIESLEIDPTDSNMHLVYGT-----GMTIFGHDLTWMD 413
Db 377 -----SSSSPG-----STVVAAPGSSS--TGGSTSPSASSSSSGTMTNSGS 416
OY 414 TRHNVISQSLADIEEFVSQDLASAPGSELLAAVGDNDG-----TFAIRNLTGTSPOV 469
Db 417 TQSTYVAVPVSSTFGSSTPIPIASSSSSGSTTVVSSGSSSTYSSSTSPSASSSAGTA--ST 474
OY 470 MATPTWATSTVDYAGNSVYRVG--NTACTOVAISSDGAATSI-----DYADTSM 522
Db 475 IGSSTGSTATIYVPGSSSSVSGSTQASPSPCGTMTSVSPSTYTVVGGSTSPAPSS 534
OY 523 NCGTVAVASADPTILWSTASSGVORSQFQGSFASVSLPAGAVIASDKKTNSEVPAAGSS 582
Db 535 PNPSSSPASTGSTITIS--GSSSIIVSVSGSTVSGSTGTSQSLASTAT-----PGSSS 588
OY 583 TF-YYSKDTGSSFTKGLGSACTIRIDIAHPPTAGTLVSTDVGFIR8--TDSGTFQ 639
Db 589 TVPSSSSPQSSQSPAPNTGSTTPTQSOTSQSPSPSNPSSSTPTGSSOSTIIPESSTASS 648
OY 640 VSTALTNTYQIAL-----GVGSGNMNLVAFGTGPGCARLYAS-----GDSGASWT 685
Db 649 PGSTGSTFVSVAETVSQSTVPSSSLGTQSTNSSPSPSLPSTSGMTSLNSEPSSST 708
OY 686 DTQGSOG-----FGSIDSTKVASGSTAGVYVGTNGCVFYAAGTVGCGTGT8- 735
Db 709 QSSGASTLTPSPNPSPQSTSSLESSTSGATTSAGAGTTMTSPSQS--SSVSSQGSTSP 767
OY 736 -----SSTKQSSSTSSASSSTTLRNVSTTRASTVTSRTSSAAG 782
Db 768 AASTTSGEMTSGCGTQPGSSVSTSAHILITQOSVSTNSPGSTYTRPSTVSGSTSSG 825

```

RESULT 15

```

J00188
ice nucleation protein - Erwinia herbicola
C:Species: Erwinia herbicola
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 26-Aug-1999
C:Accession: J00188; S35273
R:Warren, G.; Coriolo, L.
Gene 85, 239-242, 1989
A:Title: The consensus sequence of ice nucleation proteins from Erwinia herbicola, ps
A:Reference number: J00188; MUID:90152370; PMID:2515997
A:Accession: J00188
A:Molecule type: DNA
A:Residues: 1-1258 <MAR>
A:Cross-references: EMBL:M26382; NID:g148419; PIDN:AAA24823.1; PID:g148420
A:Experimental source: strain M1
R:Gurian-Sherman, D.; Lindow, S.E.; Panopoulos, N.J.
Mol. Microbiol. 9, 383-391, 1993
A:Title: Isolation and characterization of hydroxylamine-induced mutations in the Erw
A:Reference number: S35273; MUID:94018633; PMID:8412688
A:Accession: S35273
A:Status: preliminary
A:Molecule type: protein
A:Residues: 387-536 <GUR>
C:Genetics:
A:Gene: iceE
A:Superfamily: ice nucleation protein

```

```

Query Match 5.6%; Score 243.5; DB 2; Length 1258;
Best Local Similarity 24.3%; Pred. No. 2.7e-05;
Matches 187; Conservative 84; Mismatches 303; Indels 197; Gaps 34;

```

```

OY 103 GATWSTNLPFYVGGMPPRGARLAVDPANSNIIFGARGNGMLKMTDGGVYFSSKS 162
Db 165 GSTLSGTHOSQLIAGGSTETAGD-----SSTL-----AGGSTGTAGADSTLVA 210
OY 163 SFTATGTYIPDSDSDSNGVNSDKQGLMW--VTFDSTSTTGATSRIFVGTADNITASVYV 220

```

```
Db 211 GYGSTQAGESSQMGAGYSTQTMKSGDLTACYGSTGTAGDSSSLAGYGTQTAGEDS 270
QY 221 STNAGSTWMAVPGQPGYFPHKAKLOPAEKALFTY---SDGTPYDGTGSGVWRDIAG 277
Db 271 SLTAG-----YGSTQAKGSDLTAGYGTGTAGADSSL-----IAG 307
QY 278 -----GTWKDITPVSGSDLYFGFGLGLDLQKPGTAVVASLNSMWPDAOL 322
Db 308 YGSTQAGEESTQAGYGTQAKGSDLTAGYGTG-----TAGDSSL 352
QY 323 ---FRSTDSCTTSPITWAMASYPTETYYYSISTPKAPWIKNNFIDVASESPSDLIKRLG 379
Db 353 IAGYGTQTAGEDSSLTAG-----YGSTQTAQ---KGS---DITAGYGTGTA--- 394
QY 380 WMIESLEIDPTDSNMHLYGTGWTIFGGHDLTMDTRHNVSIOSLADGIEEFSVODLASAP 439
Db 395 -----GADSSLINGYST--QTAGEESTQAGYGT---STQ---TAQ 427
QY 440 GGSSELLAAGD-----DNGFTFASRNDLGTSPQTVWATPTWATSTSYDYAGNSVKSVMRV 494
Db 428 KGSDDLTAAGYSTQAGDSSSLINGYSTQTAGEDSSLTAGYGT--STQAKGSDLTAGYGS 486
QY 495 GNTAGTQVAISSDGGATWSIDVAD--TSMNGTVAAYSADGDT--LWSTASSGVQRSQEQ 551
Db 487 TSTAGYESSLLIAGYGTQTAGYGTSLTAGYGTQTAQNESDLITGYGSTTAGAN----- 541
QY 552 GSPASVSLPAGAVIASDKKTNSVFYAGSGSTFYVSKNDGSSFTRGPKLGSAGTIND--- 608
Db 542 -----SSLAGYGTQTAGYNSVLTAGYGT--QTAREGSDLTAG--YGSTTAGSDSS 591
QY 609 -IAHPPTAGTLVSTDVGIFRSTDSG-----TTFGVSTALTNVYOALGVSGSNW 660
Db 592 ILAGYGTQTAGYHSSLTAGYGTQTAGESVLTGTGYSTAGADSSLIA--GYGS---- 646
QY 661 NLVAFGTGPGSARLVASGDSGASWTDIQSGQ---GFGS-----IDSKVAGSGS--TAGQ 710
Db 647 -----TOTAGYNSILTAGYGTQTAGESGDLTAGYGTSTAGADSSSLIAGYGTQTAGY 700
QY 711 VVYGTNGRGV-----FYAOGTYG-----GGTGTSSSKKSSSTSSASSS 751
Db 701 NSILTAGYGTQTAGESGDLTSGYGTSTAGADSSSLIAGYGTQTAGYHSSLTAGYGTQ 760
QY 752 TILRSSVSTTRASTVTSRTSS--AAGPTGSGVAG--HYAOCGGIGMTGPTQ 800
Db 761 TAREQSVLTGTGYGTSTAGADSSSLIAGYGTQTAGYHSSLTAGYGTQTAQ 811
```

Search completed: April 26, 2003, 13:38:30
Job time : 75.7288 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:26:23 ; Search time 31.2736 seconds

(without alignments)
1084.865 Million cell updates/sec

Title: US-10-026-994-2

Perfect score: 4343

Sequence: 1 AFSMKNVKLGSGGFGFVPGII.....TCGVAPYCKQNDYXYQCV 818

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	264.5	6.1	537	1	GUX1_PENJA
2	257.5	5.9	1567	1	ICEN_XANCT
3	244	5.6	540	1	GUX1_ASFAC
4	243.5	5.6	1258	1	ICEN_ERWHE
5	236	5.4	1322	1	ICEN_PANAN
6	232	5.3	1150	1	APMU_PIG
7	230	5.3	1210	1	ICEN_PSEPL
8	226	5.2	516	1	GUX1_NEUCR
9	225.5	5.2	1200	1	ICEN_PSEEX
10	225	5.2	1148	1	ICEN_PSEEX
11	222	5.1	1196	1	ICEN_PSEEX
12	217	5.0	1034	1	ICEN_PANAN
13	216.5	5.0	827	1	XANP_XANS2
14	209.5	4.8	2124	1	PGCA_RAT
15	203.5	4.7	1387	1	TROP_HUMAN
16	202.5	4.7	1378	1	YSE89_CABEL
17	202	4.6	1025	1	SLAP_CAUCR
18	201	4.6	525	1	GUX1_HUMGR
19	198.5	4.6	1090	1	GUXB_CELRT
20	198.5	4.6	2132	1	PGCA_MOUSE
21	196	4.5	513	1	GUX1_TRIFE
22	195	4.5	320	1	CELI_AGAB1
23	194.5	4.5	463	1	GUX1_TRIFO
24	194	4.5	400	1	PTOA_DICDI
25	194	4.5	1140	1	YV96_YEAST
26	193.5	4.5	2090	1	N214_HUMAN
27	193	4.4	1861	1	APU_THEATV
28	191	4.4	513	1	GUX1_TRIVI
29	190.5	4.4	1325	1	YDEK_ECOLI
30	190	4.4	1654	1	OMPB_RICRI
31	189.5	4.4	1468	1	N153_RAT
32	187	4.3	962	1	GUNA_PSEPL
33	187	4.3	1075	1	F105_YEAST

34	187	4.3	2660	1	YEBJ_ECO57	Q084v7 escherichia
35	186	4.3	344	1	GUNA_TRIRE	014405 trichoderma
36	185	4.3	1609	1	FIG2_YEAST	P25653 saccharomyc
37	184	4.2	881	1	VJH8_YEAST	P47033 saccharomyc
38	183	4.2	914	1	MA22_MYCTU	006794 mycobacteri
39	183	4.2	1306	1	MSB2_YEAST	P32334 mycobacteri
40	182.5	4.2	966	1	F1B1_PETMA	P02674 petromycon
41	182	4.2	670	1	VG50_HSV1	000130 ictaluriid h
42	181.5	4.2	801	1	V747_MYCTU	053810 mycobacteri
43	181	4.2	516	1	GUX1_PHACH	P13860 phanerocha
44	180.5	4.2	459	1	GUX1_TRIFE	P07981 trichoderma
45	180	4.1	937	1	HVRI_CANAL	P46591 candida alb

ALIGNMENTS

RESULT 1
ID GUX1_PENJA STANDARD: PRT; 537 AA.
AC 006886;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulohydrolase I)
DE (1.4-beta-cellobiohydrolase).
GN CBH1.
OS Penicillium janthinellum (Penicillium vitale).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OX NCBI_TaxID=5079;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CA1.
RX MEDLINE=93178976; PubMed=8440481;
RA Koch A., Weigel C.T.O., Schultz G.;
RT "Cloning, sequencing, and heterologous expression of a cellulase-
encoding cDNA (cbh1) from Penicillium janthinellum.";
RT Gene 124:57-65(1993).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
in cellulose and cellobiose, releasing cellobiose from the non-
reducing ends of the chains.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
or send an email to license@sdb.ch).
CC EMBL: X59054; CAA41780.1; -
CC PIR: J00150; J00150.
CC HSSP: P00725; BCRL.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001722; GH_7.
DR Pfam: PF00734; CBM_1; 1.
DR Pfam: PF00840; Glyco_hydro_7; 1.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD186135; GH_7; 1.
DR SMART: SM00236; fCBD; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
KW Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 18
FT CHAIN 19 537
FT DOMAIN 19 453
FT DOMAIN 454 477
FT DOMAIN 478 537
FT ACT_SITE 235 235
FT NUCLEOPHILE (BY SIMILARITY).
FT CELLULOSE-BINDING (BY SIMILARITY).
FT CATALYTIC LINKER.
FT EXOGLUCANASE I.
FT POTENTIAL.


```

OY 489 KSVVRVGN-----TAGTQVAISSDGGATWISIDYADTSMNG----- 525
DB 705 TQTARKGSDVATGAGSTGAGADSTLLIAGYSGTQT--SSDSSTLAGYSGTQTARKGSDV 762
OY 526 TVAVSADGFTILMSTRASSGVORSQFQSGPAAV-----SSLPACAVIASDKKT 572
DB 763 TVAGYGTGTGAGADSTLLIAGYSGTQTSQSDSLTAGYSGTQTARKGSDIYAGSTGTAGA 822
OY 573 NSVFVAGSGSTFYVSKDTSSTFRPKLGSAGTIR---DIAAHPTAGTL-----Y 620
DB 823 DSTLLIAGYSGTQTSQSD--SLTAG--YGTQTARKGSDVYTAGYSGTGTAGADSTLLISY 878
OY 621 VSTDVGFIRSTDGGT--FGQVSTALNTNTQIALGVSGSNMNLVAFCTPGSA---RLY 675
DB 879 GSTQTA---GSDSSLTAGYSGTQTARKGS-DVYTAGYGS-----TGTAGADSTLLA 924
OY 676 ASGDSGASMTDQSGSFGSISTKAVAGSGTAGOYVVTNGRQVRYAOGYVGGGTGCTS 735
DB 925 GYGSQTQTSQSDSLTAGYSGTQTAR-KGSDMTAGYSGTGTAG---ADSTLLIAGYSGTQ 978
OY 736 SSTKQSS-----SSTSSASSSTLRSSVSTTRA---STVTSR 771
DB 978 TSGSSSLTAGYSGTQTAREGSDVYTAGYSGTGTAGADSTLLIAGYSGTQTARKGSDSSLTAGY 1038
OY 772 TSSAAGPTGSGVAGHYAOCGIG 794
DB 1039 GSTQTARKGSDVYTAGYSGTGTAG 1061

```

RESULT 3

```

ID GUX1_ASAPC STANDARD: PRT; 540 AA.
AC 059843:

```

```

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I) (1,4-
  beta-cellobiohydrolase I) (Beta-glucanocellobiohydrolase I).
GN CBHI.
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=50533;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F-50;
RA Takada G., Kawaguchi T., Sumitani J., Arai M.;
RT "Cloning, nucleotide sequence, and transcriptional analysis of
  Aspergillus aculeatus No.F-50 cellobiohydrolase I (cbhi) gene.";
RL J. Ferment. Bioeng. 85:1-9(1998).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
  GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
  (1) ENDOLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
  (2) EXOCELLOBIODHROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
  FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
  (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
  SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
  in cellulose and cellobiose, releasing cellobiose from the non-
  reducing ends of the chains.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY C (FAMILY 7 OF GLYCOSYL
  HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----

```

```

DR EMBL: AB002821; BAA25183.1; -.
DR HSSP: P00725; 1A26.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR001722; GH_7.
DR Pfam: PF00840; CBM_1; 1.
DR Pfam: PF00840; Glyco_hydro_7; 1.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD186135; GH_7; 1.
DR SMART: SM00236; FCBP; 1.
DR SMART: PS00562; CBD_FUNGAL; FALSE_NEG.
DR Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; Signal.
KW CELLULOSE 1
FT SIGNAL 1
FT CHAIN 23
FT DOMAIN 23
FT DOMAIN 460
FT DOMAIN 506
FT ACT_SITE 234
FT ACT_SITE 239
FT DISULFID 513
FT DISULFID 524
FT CARBOHYD 434
FT CARBOHYD 458
SQ SEQUENCE 540 AA; 57099 MW; 002D7FD28DF194D0 CRC64;

```

Query Match 5.6%; Score 244; DB 1; Length 540;
 Best Local Similarity 23.5%; Pred. No. 1.3e-06;
 Matches 150; Conservative 69; Mismatches 245; Indels 174; Gaps 31;

```

OY 234 QPKYFPKAKIQPAEKALYLYS--DGTGPDGLGSWRYDIAGTWTKITPPVSG-SD 290
DB 24 QVGTY---TAETHPS-----LTWQTCGSGSGCTTSGV---VIDANRWVHEVGYYTN 71
OY 291 LYEGFGGLDLDQKRGTLVVASLNSWMPDOLFRSDGTTWSPITAMASPTLEYYSI 350
DB 72 CYSG-----NTW--DSSIC-STDTTCASECALLEGATYEST--YGV 106
OY 351 STPKAPWIKNNFIDVYESPSDGLIKRLQWIMLESLEIDPTDSNMHLYGTGMTIFGHDLT 410
DB 107 TISGSS-LRLNVTYASQ-----KNIGSRILYLLADSDTYETPKLFNEFF--DYDVS 156
OY 411 NMDTRHNVSIGSLA-----DGIEFVSVDLASAPGSGSELLAAGVDNGFTFASRNDLGTS 466
DB 157 NLPCCGLNGALYFVSMADGDGVSFRPTNKAGAKYGTGYCQSCOPRDLKFTDGOANIEGWEP 216
OY 467 QTVMAPTWATSTSDYXNSVSVRVGNTAGTQVAISSDGGATWISIDYADTSMNGT 526
DB 217 -----SSTDVA-----AGT-----GNHSCCPEMDIMEANSTISAF 247
OY 527 VAYSAD-----GDTILMSTRASSGVORSQFQSGPAAVSLPAGAVIASDKRTNSVFYA 578
DB 248 TAMPDSDVQQTCTGDT--CGGTYSDDTOR--YSGTCD-----PDCGDFRPFQGNINFEY- 298
OY 579 GSGSTFYVSKDMGSSFTFRPKLGSAGTIRDIANAHTTACTLYVSTDVGIFRSTDSGTTFG 638
DB 299 GCGKYVDNSK-----PFTVVTQPTFHD-----STDGCTL-- 327
OY 639 QVSTALNTNTQIALGVSGSNMNLVAFCTGPS-----GARTVAS 677
DB 328 ---TEIRRLY-VONGVVIINGSPYTTAASGNSTITESFCAEKTLELGDNTVFETHGGLSAM 383
OY 678 GDSGAS-----WTDLGSGSFGSID--STKVASGSGTAGOYVVTNGRQVRYAOG-- 725
DB 384 GDALDGMVLVSLWMDHADLMLDSDVPTTSCASSPVARAGTCPTTGNMTYEANRP 443
OY 726 ---TVGGTGTGTSSTKQSSSTSSASSSTTLRSSV-VSTRASTVYSSRTSSAAGPTG 780
DB 444 NSYVYVSNIKFCTLMTSVSGTSSGSSSSSTLTTRKASTSTSSKTTTTSKSTSSSS. 503
OY 781 SGVAGHYAOCGIGMTGPHQCAVPYVCOKONYYOCV 818
DB 504 TVNAQLYGGCGGQGWTPGPTTC-ASGTCTKONDYVSQCL 540

```

```

RESULT 4
ID ICEN_ERWHE STANDARD: PRT: 1258 AA.
AC 163239:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN ICEE.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=549;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=90152370; PubMed=2515997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
herbicola, Pseudomonas fluorescens and Pseudomonas syringae."
RL Gene 85:239-242(1989).
CC -1- FUNCTION: ICE NUCLEATION PROTEIN ENABLE BACTERIA TO NUCLEATE
CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
PERIODICITY IS SUPERIMPOSED.
CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M26382; AAA24823.1;
CC PIR: J00188; J00188.
CC DR HSSP: P06620; IINA.
CC DR InterPro: IPR000258; Ice_nucleatn.
CC DR Pfam: PF00818; Ice_nucleatn; 65.
CC DR PRINTS: PRO0327; ICENUCLEATN.
CC DR PROSITE: PS00314; ICE_NUCLEATION; 45.
CC KM Ice nucleation: Repeat: Outer membrane.
CC FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1258 AA; 125084 MW; 590E8A130077EBD4 CRC64;
Query Match 5.6%; Score 243.5; DB 1; Length 1258;
Best Local Similarity 24.3%; Pred. No. 3.8e-06;
Matches 187; Conservative 84; Mismatches 303; Indels 197; Gaps 34;
QY 103 GATWSTNLPFKYGVGMGRGAGERLAVDPANSNIIFYGARSNGGLMKSTGGGVTEFSKVS 162
DB 165 GSTLSTGHQSLIAGYGSTETAGD-----SSTLI-----AGVSGTGTAGADSLTVA 210
QY 163 SFATGTIYIDPDSNNGYNDKQGLM--VTFPSTSTTGATSRIVGTADNITASVYV 220
DB 211 GYSTGTAGESSQMGAGYGTGTGCMGSDLTJAGYSGTGTGDDSSLIAGVSGTGTAGEDS 270
QY 221 STNAGSTWAVPQPKYFPHKAKLOPAEKALVLYT---SDGTGPYDGTIGSWRWYDIAG 277
DB 271 SLTAG-----YGTGTAKGSDLTJAGYSGTGTAGADSSL-----IAG 307
QY 278 -----GTWKDITPVSGSDLYFGGGLGLDLQKPGTLLVAVSLNSMWPDAQL 322
DB 308 YGSTGTAGEESTGTAGYSGTGTAKGSDLTJAGYSGTG-----TAGDDSSL 352
QY 323 ---FRSDSTTSPITWAMASYPTETYYYSISPKAPWIKNNFIDVTSSEPSGLIKRLG 379
: 11: : 1:1: : 1:1: : 1:1: :

```

```

DB 353 IAGVSTGTAGEDSSLTAG-----YGSTGTAG-----KGS--DLTAGYSGTGA----- 394
QY 380 WMEISLEIDPTDSNNHLYGKMTIFGGHDLTMDTNNHNVSIOSLANGIEFVSQDLASAP 439
DB 395 -----GQSSLIAGYSGT--QTAGEESTGTAGG-----STG--TGA 427
QY 440 GGSSELLAAVGD-----DNGFTFASRNDLGTSPQTVATPTMATSTSDVAGNSVKSVMV 494
DB 428 KGSDDLKAGYSGTGTAGDDSSLIAGYSGTGTAGEDSSLTAGYV-STGTAKGSDLTAGYGS 486
QY 495 GNTAGTQVAIISDGCATWISIDYAAD-TSNMGCVIVASAGDIT--LWSTASSGVORSQRO 551
DB 487 TSTAGYESSLIAGYSGTGTAGYSGTGTAGYSGTGTAGYSGTGTAGYSGTGTAGYSGT 541
QY 552 GSPASVSLPAGAVIASDKKTNSEFVAGSGSPFYVSKDGSFTRPKIGSGAGTIRD--- 608
DB 542 -----SSLIAGYSGTGTAGYSGTGTAGYSGT--GTAREGSDLTAG--YGSVGTAGSDSS 591
QY 609 -IAHPTTAGTLYVSTDVGFIRSTDSG-----TFEGVSTALNTQYIALGVSGSNM 660
DB 592 ILAGYSGTGTASVHSSLTAGYSGTGTAREGSLVLTGTGSGTGTAGADSLLA-GYGS----- 646
QY 661 NLVAFGTGSGARLYASGDSGASWTIOGSO--GFGS-----IDSTKVASGS--TAGO 710
DB 647 -----TOTAGYNSILTAGYSGTGTAREGSDLTAGYSGTGTAGADSLIAGYSGTGTAGY 700
QY 711 VYVGTNGRGV-----FYAGGTGV-----GGGTGSSSTKSSSTSSASS 751
DB 701 NSLLTAGYSGTGTAREGSDLTGSGTGTAGADSLIAGYSGTGTAGYSGTGTAGYSGTGTAG 760
QY 752 TLLRSSVSTTRASTVTSRTSS-AAFGTSGVAG-HYAQCGIGMTGPQ 800
DB 761 TAREGSLVLTGTGSGTGTAGADSLIAGYSGTGTAGYSGTGTAGYSGTGTAGYSGTGTAG 811
RESULT 5
ID ICEN_PANAN STANDARD: PRT: 1322 AA.
AC P20469;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein Inaa.
GN INAA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=553;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=90092494; PubMed=2599095;
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity
RT to those of Pseudomonas species and regions required for ice
RT nucleation activity."
RL FEBS Lett. 258:297-300(1989).
CC -1- FUNCTION: ICE NUCLEATION PROTEIN ENABLE BACTERIA TO NUCLEATE
CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -1- SUBCELLULAR LOCATION: Outer membrane (by similarity).
CC -1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL: X17316; CA33194.1; -
 DR PIR: S07053; S07053.
 DR HSPSP: P06620; IINA.
 DR InterPro: IPR000258; Ice_nucleatn.
 DR Pfam: PF00818; Ice_nucleation; 69.
 DR PRINTS: PR00327; ICENUCLEATN.
 DR PROSITE: PS00314; ICE_NUCLEATION; 49.
 DR Ice_nucleation: Repeat; Outer membrane.
 KW DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
 FT SEQUENCE 1322 AA; 131094 MW; 89B0EE24AA837039 CRC64;

Query Match 5.4%; Score 236; DR 1; Length 1322;
 Best Local Similarity 25.1%; Pred. No. 1e-05;

Matches 212; Conservative 81; Mismatches 332; Indels 218; Gaps 41;

103 GATMSEFTNLPFKVGNMGRGAGERLAVDPANSNITTYGANSNGMLKMTDGGVTFYSVS 162
 165 GSTLSGTHQSOLIAIGYSTETAGD-----SSTLI-----AGYSTGTAGSDSTLVA 210
 163 SFTATGYIIPPSDSNGNSDKGLMW--VTFDSTSTTGATSRIFGTADNITASYV 220
 211 GYGSTGTAGESSQMAGYSTGTGMKSDLTAGYSTGTAGDSSLIAGYSTGTAGEDN 270
 221 STNAGSTSAVPGOPKTFPIKAKIQAPKALYLT--SGTGPYDGTLSWRYDIAG 277
 271 SLTAGYSTGTAGESSSLIAGYSTGTAKGSDLTAGYSTGTAGADSSL-----IAG 323
 278 -----GTMKDLTPVSGSDLYFGFGLG-----DLQKRGTLV 310
 324 YGSTGTAGEESTGTAGYSTGTAKGSDLTAGYSTGTAGDSSLIAGYSTGTAGE--D 381
 311 ASLNSMWPDAOL-----FRST-DSGTTWSPIMAWAS-----YPTETYYYSISTP 353
 382 SSLIAGYSTGTAKGSDLTAGYSTGTSGADSSLIAGYSTGTAGESTGTAGYG-STQ 440
 354 KAPMKNNFIVTSSPSDGLIKLGMIESELDPTDSNMHLYGTGTTTGGHDLTMMWD 413
 441 TA---OKGSDLTAGYSTGTAGDSSLIAG-----YGSTGTAKGSDLT--- 481
 414 TRHNYSIGSLADGIEFEFVODLAS---APGSELLAAVDDNGFFAFARNDLGT---SPQ 467
 482 --AGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYST 536
 468 TWMTPT-----WMTSTSVYAGNSVSVYAVGN-----TAGTOVAISSDG 508
 537 TAGANSSLIAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYST 596
 509 GATMSIDYAAPTSMNGG-----TVAYSA-----DGDITLMSTASSGVRSQFQGSFASY- 557
 597 GSTSTA--GADSSLIAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYST 651
 558 --SSLIPAGAVIASDKRTYGVYAGSGSTFYYSKIDGSSFTGPKUGSAGTIRD---IAA 611
 652 ADSSLIAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAG 707
 612 HPTTAGLYVSTDVGIFRSTDSG-----TTFQGVSTALNTYQIALGVGS---GSNM 660
 708 YGSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAG 765
 661 NLVAFGTGPS-----GARLYASGDSG-----ASTDTIOG 689
 766 SILTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAG 825
 690 SQ-----IDSTKVAGSGS--TAGOVYVGTNGRG---VEYAGQVFGGCTG 732
 826 AGRSNDLTGCGSTGTAGDSSLIAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAG 885
 733 GTSSTKQSS-----STSSASSSTTLRSVSVT---TRASTVTSRTSSAAGPTGSGV 783
 886 STTAGYDSSSLIAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYSTGTAGYST 945
 784 AGH 786
 111

DB 946 AGY 948
 RESULT 6
 APMU_PIG STANDARD: PRT: 1150 AA.
 AC P12021;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Apomucin (Mucin core protein) (Fragment).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 NCBI_Taxid=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Submaxillary gland;
 RX MEDLINE=91236743; PubMed=2033060;
 RA Eckhardt A.E., Timpl C.S., Abernethy J.L., Zhao Y., Hill R.L.;
 RT "Porcine submaxillary mucin contains a cysteine-rich,
 RT carboxyl-terminal domain in addition to a highly repetitive,
 RT glycosylated domain";
 RL J. Biol. Chem. 266:9678-9686(1991).
 RN [2]
 RP SEQUENCE OF 1-503 FROM N.A.
 RC TISSUE-Submaxillary gland;
 RX MEDLINE=88087170; PubMed=2826455;
 RA Timpl C.S., Eckhardt A.E., Abernethy J.L., Hill R.L.;
 RT "Porcine submaxillary gland apomucin contains tandemly repeated,
 RT identical sequences of 81 residues";
 RL J. Biol. Chem. 263:1081-1088(1988).
 RN [3]
 RP SEQUENCE OF 45-80.
 RC TISSUE-Submaxillary gland;
 RX MEDLINE=87280230; PubMed=3611111;
 RA Eckhardt A.E., Timpl C.S., Abernethy J.L., Toumadje A.,
 RA Johnson W.C. Jr., Hill R.L.;
 RT "Structural properties of porcine submaxillary gland apomucin";
 RL J. Biol. Chem. 262:11339-11344(1987).
 RN [4]
 RP CARBOHYDRATE-LINKAGE SITES, AND SEQUENCE OF 45-125.
 RC TISSUE-Submaxillary gland;
 RX MEDLINE=97248516; PubMed=9092502;
 RA Gerken T.A., Owens C.L., Pasumathy M.;
 RT "Determination of the site-specific O-glycosylation pattern of the
 RT porcine submaxillary mucin tandem repeat glycopeptide. Model proposed
 RT for the polypeptide:galnac transferase peptide binding site.";
 RL J. Biol. Chem. 272:9709-9719(1997).
 CC -1- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN
 CC SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY
 CC GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE
 CC THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL
 CC ENVIRONMENT.
 CC -1- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A
 CC MULTIMERIC MUCIN STRUCTURE.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
 CC -1- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81
 CC RESIDUES.
 CC -1- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF
 CC THE REPEAT UNITS. HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER
 CC RESIDUES WHICH HAVE GLY AT POSITIONS AT +2 OR -2 FROM THE
 CC GLYCOSYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE, THE
 CC PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO
 CC ENHANCE GLYCOSYLATION.
 CC -1- SIMILARITY: CONTAINS 1 WFEC DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE (CTCK) DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

DR InterPro: IPR001722; GH_7.
DR Pfam: PF00734; CBM_1; 1.
DR Pfam: PF00840; Glyco_hydro.7; 1.
DR ProDom: PD001821; CBD_fungal; 1.
DR ProDom: PD186135; GH_7; 1.
DR SMART: SM00236; fCBD; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 17
FT CHAIN 18 516
FT DOMAIN 18 445
FT DOMAIN 446 480
FT DOMAIN 481 516
FT ACT_SITE 223 223
FT ACT_SITE 228 228
FT DISULFID 488 505
FT DISULFID 499 515
SQ SEQUENCE 516 AA; 54471 MW; 38E598406EA81900 CRC64;

Query Match 5.2%; Score 226; DB 1; Length 516;
Best Local Similarity 26.1%; Pred. No. 1.2e-05;
Matches 78; Conservative 37; Mismatches 88; Indels 96; Gaps 11;

OY 563 GAVIASDKNTSYFYAGSGSTFVYSGDTGSSFTRGPKL--GSAGTIRDIAPHPTTAGTLY 620
DB 271 GCDPNSYRNGNTTFY--GEGKTV---DTSKFTVVTQFIKDSIGDLAEIKAFYVQNGKVI 325
OY 621 VSTDVGIFRSTDSG-----TTEG-----QVSTALNTYQIALGVSGS 658
DB 326 ENSQSNVDGSGNSITOSPFCKSKTAFGDIDNNKKGGLKQKALQANVLMSTI---- 381
OY 659 MNMLVAF-----GTGPGSARLYASGDSGASWTDIOGSGFGSIDST 699
DB 382 -WDDHANMLMLDSTYVPKVPAGYRSGGPTTSGVPAEVDANA-----PNS 426
OY 700 KVGSGSTAGOVYVNGRGVFAAGTGVGGTGTGSSSTTQSSSSSTTIRSSV 759
DB 427 KAFPSIKRCHLGI-----SPFGSGSGTTPPSNPSSASPTIS----- 463
OY 760 STTRASTVTSRTSSAAGPTGSGVAGHYAOCGIGMTGTQCVAPYQKONDYVYOCV 818
DB 464 -----STAKRSSTSTASNPSCGTG-AAHMAQCGIGSGPTTCPEPTCANDHDYISQCV 516

RESULT 9
ICEN_PSES 9
ID ICEN_PSES 9 STANDARD: PRT: 1200 AA.
AC P06620:
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
DE INAZ.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=321;
RN 11
RN SEQUENCE FROM N.A.
RC STRAIN=5203;
RA Green R.L., Warren G.J.;
RT "Physical and functional repetition in a bacterial ice nucleation
RT gene.";
RL Nature 317:645-648(1985).
RN 12
RN SUBCELLULAR LOCATION.
RX MEDLINE=92297969; PubMed=2520825;
RA Lindov S.E., Lahue E., Govindarajan A.G., Panopoulos N.J., Giles D.;
RT "Localization of ice nucleation activity and the iced gene product in
RL Pseudomonas syringae and Escherichia coli.";
RL Mol. Plant Microbe Interact. 2:262-272(1989).
RN 13
RN 3D-STRUCTURE MODELING OF 490-535.
RX MEDLINE=93360260; PubMed=8355267;

RA Kajava A.V., Lindov S.E.;
RT "A model of the three-dimensional structure of ice nucleation
RT proteins.";
RL J. Mol. Biol. 232:709-717(1993).
CC - FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC CRYSTALLIZATION IN SUPERCOOLED WATER.
CC - SUBCELLULAR LOCATION: Outer membrane.
CC - DOMAIN: CONTAINS 122 IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDE AND A
CC REGIONAL 48-RESIDE PERIODICITY IS SUPERIMPOSED.
CC - MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC - SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X03035; CA26837.1; -
DR PIR: A24405; A24405.
DR PDB: 1INA; 31-OCT-93.
DR PDB: 1INB; 31-OCT-93.
DR InterPro: IPR000258; Ice_nucleatn.
DR Pfam: PF00818; Ice_nucleation; 61.
DR PRINTS: PR00327; ICENUCLEATN.
DR PROSITE: PS00314; ICE_NUCLEATION; 40.
KW Ice nucleation; Repeat; Outer membrane; 3D-structure.
FT DOMAIN 176 1151 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1200 AA; 118587 MW; E3BEAD4BEA67AB80 CRC64;

Query Match 5.2%; Score 225.5; DB 1; Length 1200;
Best Local Similarity 24.3%; Pred. No. 3.5e-05;
Matches 198; Conservative 88; Mismatches 292; Indels 237; Gaps 39;

OY 91 PSNGAIIRSSDRCATNSFTNLPRKVGNNMPCRGAGERLAVDPANSNITY-FGA--RSNG 147
DB 170 PQN---VQRTATYSTLSGDMHSRLIAGYGSNETAGN-----HSDLIAGYSGTGA 219
OY 148 LW-----KSTDG-----VTFKVSSTFA---TGTYPDPSPDSNGYSDKO 185
DB 220 SMLVAGYGTQTAGDSALTAGYGTQTAGESNLTAGYGTAGSDSSLIAGYSTO- 278
OY 186 GLMWYFEDSTSSSTTGATSLIPVGTADNTASVYVSTNAGSTSNAPGQ-----PK 237
DB 279 ----TSGGSSLTAGYGTQTAGESNLTAGYGTGTAGSDSSLIAGYGTQTAGSDSS 333
OY 238 YPFHKKALQPAERKALVLTYY---SDGTGPYDGTGYSVWRVYDIAG-GTWKDDITPVSGSD--L 291
DB 334 LTAGYGTQTAGESNLTAGYGTAGYVDSL-----IAYGS-----TQTSGSDSL 382
OY 292 YFEGGSLGLDQKPGTLVVASLNSWMPDAQLFIRSTSGTTWSPIMAWASYPTETYYYSIS 351
DB 383 TAGYGT--STQTAGESNLTAGYGS-----TGTAGSDSSLIAGYGTQTAGSDS 428
OY 352 TRKAPMINKNFIDVTSESPESDGLIKRLGMIESLEIDPTDSNIMVIGTMTTCGHDLTN 411
DB 429 SLTAGY-----GSTQTAGESNLTAGY--GSTGAGYVDS--LIAGYGTQTAGSD-- 475
OY 412 WDTRHNVSIOGLADGLEEFVSODLASAPGSELLAAGVDDNGFTFASRNDLGTSPQTVMA 471
DB 476 -----SALTAGYGT--STQ--TAGESNLTAGY----- 499
OY 472 TPTWATSTVDYAGNSVYRVNGTAGTOVAISSDGAWTMSIDYADTSMNGTV--AY 529
DB 500 -----STGTAGADSSLIAGYGTQTAGESNLTAGYGT-----QTRREGSTLTAGY 546
OY 530 SADGDTITLMSTASSGVORSPFGSGFASVSSLPAGAVIASDKNTSYFYAGSGST----- 583

Db 547 GGTGAGDSSLIAGYSTGTGSG---SESSLIAGYSTGTGTAOAGSVLTGSGYSTGTAGAA 603
 QY 584 -----FVYSKDTGSSFTGRPK-----IGSAGTIRD-----IAAHTTAG 617
 Db 604 SMLTGTGYSTGAGHESITLIGYSTGTGAKHSILTAGYSGTGTARDSDLIAGYST-G 662
 QY 618 TL-----YVSTDVGIFRS-----TDSGTFPGOVSTALNTYQIALG 653
 Db 663 TAGSGSSLIAGYSTGTASYSRMLTAGYSGTGTAREHSDLVLTGSGYSTAGSNSLLA-G 721
 QY 654 VGSNGMNLAFGCTGPGSARILASDSCASWTDIGSQ-----GFGS-----IDSTVAGSG 705
 Db 722 YGS-----TGTAFKSLITAGYSGTGTAEERTSLVAGYSGYSTAGYSSSLIAGY 771
 QY 706 STAGOVYGTNGRGVFAAG-----TVGGG---TGGSSTSKSSSTSSASSSTLR 755
 Db 772 STGTAGYSTLTAGYSGTGTAEENSLLTGTGYSTAGYSSSLIAGYSTGTAGYESTLT 831
 QY 756 SSVVST---TRASTVTSRTSSAAGPTGSGVAGH 786
 Db 832 AGYGTGTAEERSDLVTGSGYSTAGYSSSLIAGY 866

RESULT 10
 ICEK_PSESX STANDARD: PRT: 1148 AA.
 AC 030611:

DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN INAK.

OS Pseudomonas syringae.
 CC Bacteria: Proteobacteria; gamma subdivision: Pseudomonadaceae;
 CC Pseudomonas.

OK NCBI_TaxID=317;

RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-KCTC 1832;
 RA Jung H.-C., Pan J.-G.;

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

CC -I- FUNCTION: ICE NUCLEATION PROTEIN ENABLE BACTERIA TO NUCLEATE
 CC CRYSTALLIZATION IN SUPERCOOLED WATER.

CC -I- SUBCELLULAR LOCATION: Outer membrane (By similarity).

CC -I- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
 CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A

CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.

CC -I- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
 CC FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).

CC -----

DR EMBL: AF013159; AAB66891.1; -

DR HSSP: P06620; IINA.

DR InterPro: IPR000258; Ice_nucleatn.

DR Pfam: PF00818; Ice_nucleatn; 57.

DR PRINTS: PRO0327; ICENUCLEATN.

DR PROSITE: PS00314; ICE_NUCLEATION; 38.

KW Ice nucleation; Repeat; Outer membrane.

FT DOMAIN 161 164 POLY-ALA.

FT DOMAIN 180 1099 OCTAPEPTIDE PERIODICITY.

SO SEQUENCE 1148 AA; 113811 MW; BAA019CF20FAE224 CRC64;

Query Match 5.2%; Score 225; DB 1: Length 1148;
 Best Local Similarity 23.5%; Pred. No. 3.5e-05;
 Matches 182; Conservative 79; Mismatches 286; Indels 228; Gaps 32;

QY 111 LPEKVGNMPPGRGACERLAVDPANSNIIFYGARGNGIMLKSGDGVTEKSSFPATGTY 170
 Db 169 LPVTTPOMLQTATYGTSTLSDGNHSLIAGYSGNENAGNHSIDLIAGYSGTGTAGYSTQTS 228
 QY 171 IPDPDSKNGYNSDKOGKLMKWTVPDSTSTTGATSKIFVGTADNITASYVSTNAGSTWSA 230
 Db 229 GHDSSLTAGYSGSTQ-----TAQEGSNLTAGYSGTGTAGSDSSLIAGYSGTQSGSDSL 282
 QY 221 VPGQGVKTFPHAKLQAPAEKALILTY-----SDGTGPIYDGLGSLVMRYDIAG-CTMKMDIPV 286
 Db 283 TAGY-----GSTGTAEAGSNLTAGYSGTGTAGVDSL-----TAGYGS---TGT 323
 QY 287 SGSD-LYFEGGLGLDLQKPGTLVVASLNSWMPDAQLFRSTSGTWSPIAMASPYE 344
 Db 324 SGSDSALTAGY--STGTAEAGSNLTAGYSG-----TGTAGSDSSLIAGYGS- 368
 QY 345 TYYSISTPKAKWIKNFDVTSSEPSDGLKRLGMIESLEIDPTDSNHWLYGTGMTIF 404
 Db 369 -----TQTSQSD-----SLTAG-----YGSTGTQAO 389
 QY 405 GGHDLT-NMDFTHNYSIOS---LADGIEFVQDLASAP-----GSELLAAVGD 451
 Db 390 EGSNLTAGYSGTGTAGVDSLIAGYSGTQSGDSALTAGYSGTGTAEAGSNLTAGYSG- 447
 QY 452 NGFTFASRNDLGTSPQFWATPTMATSTSVYDAGNSVKSVAEVTAGTQVAISSDGAT 511
 Db 448 -----STGTAGADSSLIAGYSGTQSGESSLTAGYSGT 481
 QY 512 WSIDYADTSMNGTV--AVSADGDTILMSTASSGVQSRQFQGSFASVSLPAGAVIASD 569
 Db 482 -----GTAREGSTLTAGYSGTGTAGADSSLIAGYSGTQSG---SESSLIAGYSGTGT 531
 QY 570 KKTNSVFYAGSGST-----FVYSKDTGSSFTGRPK-----LG 601
 Db 532 AOGGSVLTSGYSTGTAGASNLTTGYSGTAGHESFTIAGYSGTGTAGKHSILTAGY 591
 QY 602 SAGTIRD-----IAAHTTAGTL-----YSTDVGIFRS-----TDS 633
 Db 592 STGTARDGSLIAGYSGTGTAGSSSLIAGYSGTGTASYSRMLTAGYSGTGTAREHSDL 650
 QY 634 GTTFCOVSTALNTYQIALGVGSGSNMNLAFGCTGPGSARILASDSCASWTDIGSQ-- 691
 Db 651 VYGYSTAGSNSLLA-GYGS-----TGTAGFKSLITAGYSGTGTAEERSDLV 699
 QY 692 -GFGS-----IDSTVAGSGTAGOVYGTNGRGVFAAG-----TVGGG---TGGS 735
 Db 700 AGYGTSTAGYSSSLIAGYSTGTAGYESTLTAGYSGTGTAEENSLLTGTGYSTAGIS 759

RESULT 11
 ICEV_PSESX STANDARD: PRT: 1196 AA.
 AC 033479:

DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN INAV.

OS Pseudomonas syringae.
 CC Bacteria: Proteobacteria; gamma subdivision: Pseudomonadaceae;
 CC Pseudomonas.

OK NCBI_TaxID=317;

RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-INA5;
 RA MEDLINE-97462815; PubMed-9323042;
 RA Schmidt D., Pridmore D., Capitani G., Battistutta R., Neeser J.-R.,
 Janin A.;

RT "Molecular organisation of the ice nucleation protein Inav from


```

OY 123 GAGERLAADVPANSNIIFYGARGSGNGLMKSTGDGV-----TFKVSSEFTA-----TGTYIPDP 174
DB 243 GVGSTGTAGDSSLLAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAGYSTGTAGDSS 302
OY 175 SDSNGYNSDKQGLMWTFDSTSTTGATSRIFVGTADNITASVYSTNA----- 224
DB 303 SLIAGYGSQTQ-----TAGESTGTAGYSTGTAGDSSLLTAG-YGSTGTAGDSSLLTAG 355
OY 225 -GSTSAVPAGOGKFFPHAKLOPAEKALYLYL-----SOSTGTYDGTLSGVMWYDIAG--- 277
DB 356 YGISTQTA--GEDESSLTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAG-----TAGYGS 406
OY 278 -----GTWKDITPVSGSDLYFEGGGLGLDLOKPGTLYL-VASLNSWMPAQL-- 322
DB 407 TOTAGEESTGTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAG 466
OY 323 -FRSTDSGTTMSPYVA-WASYPTETYYYSI-----STPAPIKKNFIDVTSESPDGLI 375
DB 467 GYGSTGTAGDSSLLTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAG----- 506
OY 376 KRLGMMIESLEIDPDSNIMWLTGTMITFGHDLTNMOTRHNVSIOSLADGIEEFSVODL 435
DB 507 -----GYGSTLTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAG----- 537
OY 436 ASAPGSELLAAGVGDNDNGFTFASRNDLGTSPQVWATPTMATSTVDYAGNSKSVYRVG 495
DB 538 AGA--NSSLIAGY-----STGT-----ASYNSLTGAGYSTGTAGDSSLLTAG 573
OY 496 N--TGA-----ISSDGAATWSIDYADTSMNGTVAYSADG-DTTL--WSTA 541
DB 574 SULTFAGYSGSTGTAGDSSLLTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAG 631
OY 542 SSGVORSGOGSEFASY-----SILPAGAVIASDKKTNSVFVAGSSSTFYVSKDNCSSST 595
DB 632 QTAQERSDITTYGTSTAGADSSLLTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAG 689
OY 596 RQPKLGSAGTITDIAHPTTACTLYSTVDYGFIRST-----DSGTFEGOVST 642
DB 690 TG--YGSST-----AGYESSLIAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAG 743
OY 643 ALTNTRYQALGVGSSNMNLVAFGTGPGSARLYASGDSGASWTDIQSGGFGSIDSTKVA 702
DB 744 SMAG--YDSSLLTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAG 802
OY 703 GSGS--TAGQVYVGTNGRVFAAG--TVGGGTGTSSTQSSSSTSSASS--TTLRS 756
DB 803 GYGSSLTGINSRFLTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAGYSTGTAGDSSLLTAG 862
OY 757 SVVSTTRASTVYSSRTSSAAGPTGSGVAGH 786
DB 863 SLIAGPESTQITGNRSMLIAGKSSOTAGY 892

```

```

RN [2]
RP MUTAGENESIS OF ASPARTIC ACID AND GLUTAMIC ACID RESIDUES, AND ACTIVE
RX SITES.
RA MEDLINE-99419069; PubMed-10486127;
RA Oyama H., Abe S.-I., Ushiyama S., Takahashi S., Oda K.:
RT "Identification of catalytic residues of pepstatin-insensitive
RL carboxyl proteinases from prokaryotes by site-directed mutagenesis.";
CC J. Biol. Chem. 274:27815-27822(1999).
CC -1- CATALYTIC ACTIVITY: Cleavage of casein.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: AUTOCATALYTICALLY PROCESSED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S53.
CC -1- SIMILARITY: CONTAINS 1 PKD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D83740; BAA12093.1; -.
CC DR MEROPS: S53.002; -.
CC DR InterPro: IPR000601; PKD_domain.
CC DR Pfam: PF00801; PKD: 1.
CC DR SMART: SM00089; PKD: 1.
CC DR PROSITE: PS50093; PKD: 1.
CC KM Hydrolase; Protease; Serine protease; Zymogen; Signal.
CC FT SIGNAL: 1 23
CC FT PROPEP: 24 237
CC FT CHAIN: 238 635
CC FT PROPEP: 636 827
CC FT DOMAIN: 635 722
CC FT ACT_SITE: 406 406
CC FT ACT_SITE: 544 544
CC FT ACT_SITE: 585 585
CC FT ACT_SITE: 585 585
CC SQ SEQUENCE 827 AA; 83706 MW; 21A33C4C683BDC8F CRC64;

```

Query Match 5.0%; Score 216.5; DB 1; Length 827;

Best Local Similarity 21.7%; Pred. No. 7e-05; Matches 189; Conservative 75; Mismatches 291; Indels 317; Gaps 39;

```

OY 41 LNADDSWTAIVTGIADNA-----GWNHNGIDAVNALDPD--DQKYVAAGMTTNSMDPSN 93
DB 156 ISADGTAGATNGFRTSIRKFSANGREFPANDAPALVPASLDSVAVALGLONVSVKHTL 215
OY 94 GAIIRSSDRGATWSPFTNLPFKYGNMPCRGAGERLAVDPANSNIIFYG--ARSGNGLM 149
DB 216 HHVYHEDE-----VTVF--GPNVGTQAAVAANHHQDFRAATYGGSSLPAAATTTAVG 265
OY 150 KSTDGCV--TFKVSSEFTA-----TGTYIPDPSDSNG-YNSDKQGLMW- 190
DB 266 IITWGSITQTVDLNSFTSGAGLATVNSTITRVGSGTFANDP--DSNGEWSLDSODIVGTA 324
OY 191 -----TTPSTSS--TTGATSRIFVGTADNITASVYSTNA--SAYVPGOG 236
DB 325 GGVKQILFTYSANGDSSSGITDAGITASYNRAVYDNIKLINVSIGEDETAQOOSG--- 381
OY 237 KYFPHAKILOPAEKALY-----LTYSDDTG-----PYDGTLSVWRKYDIAGTWK- 281
DB 382 -----TQAADDAIFQQAIVAGOGTFSIASGDAGVYQWSTPTSSPEYVANSAGTVKI 433
OY 282 DITPVVS--GSDLYFGGGLGLDLOKPGTLYVASLNSWMPAQLFRSTDSGTTMSPYVA 337
DB 434 DLTHVSVSEPPASPYV-----IYVGGTTL-----STSGTWSGTEV 469
OY 338 WASYPTETYYYSISPPKAWIKKNFIDVTSSEPSOGLIKRLGMMIESLEIDPDSNIMWLY 397
DB 470 KNE-----GLSAIAPSOQ-----DNNQRIWM 489
OY 398 GTGWTIFGCHDLTNMOTRHNVSIOSLADGIEEFSVODLASAPGSELLAAGVGDNDNGFTFA 457

```

```

Db 490 ATG-----GVSLEY-----499
Oy 458 SHNDLSTSPQWATPTWATSTVDYAGNSKVVVNGTACTQVAISSDGGATMSIDYA 517
Db 500 -----MAPSMQSSVS-----SSTRKVGDDLAFDAASS--GALIYV--534
Oy 518 ADTSMNGTVAVSADDDTLWSTASSGVORSPQGSFASVSSLPAGAVIASDKTNSVFY 577
Db 535 GSTEVOGCT-----SLAS-----PLFVGAFARIESANNAI-----565
Oy 578 AASSGTFYVSKDGSSEFTGPKLGSAGTIRDAIAHPTTGLTYVSDVGIFFSTDSCTFE 637
Db 566 GPPASFTFOAPFTQTSLLHDVTSNGNG-----YOSHGYTATG-----DEATGE 610
Oy 638 GOVSTALTWTYQIA---LGVSGSN-----WNLYAFGTSPSGARLYASGDSGAS---M 684
Db 611 GSPFDGLKLTVAQAMNVVGGGGSTNAPPVANSVATGLVATFTDSSDGSISAHM 670
Oy 685 PDIQGSQCGSIDSTKVACSGSTAGOVYVGTNGRGVFAQGYGGCTGCTSSSTKO-SSS 743
Db 671 T-----FGD-----GSTSTATSPSHTYSAAGTVSAEFTVDNAGATSTKTSSTVTS 716
Oy 744 SRTSSASSSTTLRSVYSTRASVYTSRRTSSAAGPT-----GGGVAGHYAQCGG 792
Db 717 SSGGTGGGTVLQNGVAAATGLSAKKNGQLKYTYVPSGANSKLAISGIGDADLYVKFGS 776
Oy 793 IGMTGPTQCVAPYV-----COKONDYVY 815
Db 777 APTTSTYDC-PPYVTGNTESCSPASQDTGY 807

RESULT 14
PCGA_RAT ID STANDARD: PRT: 2124 AA.
AC P07897:
DT 01-AUG-1988 (rel. 08. Created)
DT 01-FEB-1991 (rel. 17. Last sequence update)
DT 15-JUN-2002 (rel. 41. Last annotation update)
DE AggreCAN core protein precursor (Cartilage-specific proteoglycan core
DE protein) (CSPCP).
DE AGC1 OR AGC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86087070; PubMed=3693370;
RT "Complete primary structure of the rat cartilage proteoglycan core
RT protein deduced from cDNA clones."
RL J. Biol. Chem. 262:17757-17767(1987).
RN [2]
RP REVISION TO 698.
RA Doerge K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RL J. Biol. Chem. 263:10040-10040(1988).
RN [3]
RP SEQUENCE OF 1856-2124 FROM N.A.
RX MEDLINE=86250698; PubMed=2424893;
RA Doerge K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;
RT "Partial cDNA sequence encoding a globular domain at the C terminus
RT of the rat cartilage proteoglycan."
RL J. Biol. Chem. 261:8108-8111(1986).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC similarity).
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS

```

```

CC CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC CC AND G3.
CC CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
CC CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL, M13518; AAA1836.1; -
CC CC EMBL, J03485; AAA21000.1; ALT_SEQ.
CC CC PIR, A23835; A23835.
CC CC PIR, A28452; A28452.
CC CC HSP, P98066; ITSG.
CC CC InterPro: IPR003006; Ig_MHC.
CC CC InterPro: IPR003596; Ig_V.
CC CC InterPro: IPR001304; Lectin_C.
CC CC InterPro: IPR000538; Link.
CC CC InterPro: IPR003324; SGXSG.
CC CC InterPro: IPR000436; Sushi_SCR_CCP.
CC CC Pfam, PF00047; Ig_1.
CC CC Pfam, PF00059; Lectin_c_1.
CC CC Pfam, PF00084; sushi_1.
CC CC Pfam, PF00193; Xlink_4.
CC CC Pfam, PF02339; SGXSG; 58.
CC CC PRODOM: PD000918; Link; 4.
CC CC SMART: SM00032; CCP; 1.
CC CC SMART: SM00034; CLECT; 1.
CC CC SMART: SM00406; IGV; 1.
CC CC SMART: SM00445; LINK; 4.
CC CC PROSITE: PS00290; IG_MHC; 1.
CC CC PROSITE: PS01241; LINK; 4.
CC CC PROSITE: PS00615; C-TYPE-LECTIN_1; 1.
CC CC PROSITE: PS50041; C-TYPE-LECTIN_2; 1.
CC CC KMW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
CC CC Repeat; Immunoglobulin domain.
CC CC FT SIGNAL 1 19
CC CC FT CHAIN 20 2124
CC CC FT DOMAIN 44 140
CC CC FT DOMAIN 170 247
CC CC FT DOMAIN 268 349
CC CC FT DOMAIN 504 581
CC CC FT DOMAIN 602 683
CC CC FT DOMAIN 1910 2036
CC CC FT DOMAIN 2040 2098
CC CC FT DOMAIN 48 140
CC CC FT DOMAIN 152 247
CC CC FT DOMAIN 253 349
CC CC FT DOMAIN 486 580
CC CC FT DOMAIN 587 682
CC CC FT DOMAIN 685 798
CC CC FT DOMAIN 801 1226
CC CC FT DOMAIN 1227 1909
CC CC FT DOMAIN 1910 2124
CC CC FT DISULFID 51 133
CC CC FT DISULFID 175 246
CC CC FT DISULFID 199 220
CC CC FT DISULFID 273 348
CC CC FT DISULFID 297 318
CC CC FT DISULFID 509 580
CC CC FT DISULFID 533 554
CC CC FT DISULFID 607 682

```

```

FT DISULFID 631 652 BY SIMILARITY.
FT DISULFID 1914 1925 BY SIMILARITY.
FT DISULFID 1942 2034 BY SIMILARITY.
FT DISULFID 2010 2026 BY SIMILARITY.
FT DISULFID 2041 2084 BY SIMILARITY.
FT DISULFID 2070 2097 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1842 1842 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 2124 AA; 221117 MW; E30BBE61593A34B1 CRC64;

Query Match 4.8%; Score 209.5; DB 1; Length 2124;
Best Local Similarity 21.6%; Pred. No. 0.000523;
Matches 203; Conservative 107; Mismatches 323; Indels 307; Gaps 47;

QY 64 GIDAVLADPDDOKVYAAYGMYT-----NSMDPSNGAIRSSDRG--ATWSEFTNPFKY 115
DB 713 GYDAVLELEETFEVY-----FTTEPEKQTEMEPATPVGTSLPGLPTW----- 758
QY 116 GGNMPEKGERLAVDPANSNIIFGARGNGLMKSTDCGVTFKVSSTFANGTYIPDS 175
DB 759 ---LPTVPAAEHTESPASQ-----EPSAQVPTSEEPYT---PSLAVPSGTELPSSG 807
QY 176 DSNQY-----NSDKQGLMNVTFDSTSSITGATSRIFVGTADNI---TASYVST 222
DB 808 DTSAGDPLSGDTFTDTSGRLL---DSSGEPSGSESGIPSDLDSSGLGPTVSSGLPV 863
QY 223 NAGST-----WSAVPGCPGKYPHAKLQPAEKALYLYTS-DGTGPYDG--TLGSVMR 272
DB 864 ESSASAGDEIEMPSRPYTD-----RLPSGSELEGSASASGIGDLSGLPGEIITE 915
QY 273 YDIAGTMDITPVSGSDLYFGFGCL-----GLDLQKPGTLVVASLNSMPPDAQ 321
DB 916 TSASGTEELISGLPBGDDDETSTGIDGASVLPFGRGLETASAGEVEDLSGLPS----- 969
QY 322 LFRSTDSGTTWSPIMAWASPTETYYYSISTPKAPWIKNNFIDVSESPDGLIKRLGM 381
DB 970 ---GEESESTSGIEDISVLP-----TGSPETSAAGVDLSGLPSG----- 1010
QY 382 ISLEID-----PTDSNMHLVGTMTIFGHDLTNMDT--RHNV----- 418
DB 1011 -ESLETASAGEVEDVQLPTEERG---GLETSAGIEDITVLPFGRENLETASAGEVEDVS 1065
QY 419 -----SISLADGIEFS-----VQDLASA 438
DB 1066 LPSGKBLGLETASGIEDISVFPTAEGLTASAGGVSGIPSGEDGTETSTSGVEGVSGL 1125
QY 439 PGSESL-----AAGDNGFTFASRNDLGTSPQVWAT--PTMA--TSTSDVANGSVKSY 491
DB 1126 PSCGELLETASAGEVDLG--LPTROLETASAGVDVTCIPSGREDTETSYPGVDDLSGL 1183
QY 492 VRYNAGTQVALISD-----GATWSIDYADTSMNGSTVYASADGDTILMS 539
DB 1184 PSCGELLETASAGAEIDLGLIPSGEKEDLVGASAGALDFG--KLPSGTL---GSGGT--- 1233
QY 540 TASSVQROKFOGSPASV-----SSLRAG--AVIASDKTKTSVFAAGSGS-- 562
DB 1234 PEASGLP--GSGSEYSGVIGSGPSSGLDPDFSLPGFPTVSLVDSTLVEVITATVASEL 1292
QY 583 ---TFYVKNDCSSFTPRKLGASCTIRIDIAHPT-----TAGTLVYSTVVGJFRST 631
DB 1293 EGRGITSV---GSGRESGPPLELSDSADISGLPSCGTELSGTSGLVDYSGTSGFFDY 1349
QY 632 DSGTTEGQVST-----ALTNTYQALGVGSGSNMNLVA 664
DB 1350 -SGQPGSSGTCGEGTSGIPEVSGQAVRPPDTEIISLSSGQPPVSGSGS-----IL 1403
QY 665 FCGPGSGARLYASGD--SGASWTDIOGS--QGFSGIDST-----KVAGSGTAQVY 712

```

```

DB 1404 FSGGSSGITSVSGETSGIS--DLGQPSGFPVLSCTPCTPDLASGANGSGDSSGTF 1461
QY 713 VGN-----*-----GRGVFAOCTVGGR--GGTSSSTKSSSTSSASS----- 751
DB 1462 VDTLSLEVPPTTFREEGLGASVELSGLPGEFDLSSTGCMVDVSGGSAIDSSGLISPT 1521
QY 752 ---TTLRSSVSTT--RASVTSSRTSSAAGPTGSGVAGHY 787
DB 1522 PERSGLPSCVAEVSCEVSGEVGSSLSGCAFDGSGLVSGF 1561

RESULT 15
TROPO-HUMAN
ID TROP-HUMAN STANDARD: PRT: 1387 AA.
AC 012816; 09NU89; 09UPN8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trophoblast.
TRO OR KIA1114.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Ambrose K.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 639-1387 FROM N.A.
RX MEDLINE=95278733; PubMed=7758945;
RA Fukuda M.N., Sato T., Nakayama J., Klier G., Mikami M., Aoki D.,
RA Nozawa S.;
RT "Trophoblast and tasin, a novel cell adhesion molecule complex with
RT potential involvement in embryo implantation.";
RL Genes Dev. 9:1199-1210(1995).
CC -!- FUNCTION: COULD BE INVOLVED WITH BYSTIN AND TASTIN IN A CELL
CC ADHESION MOLECULE COMPLEX THAT MEDIATES AN INITIAL ATTACHMENT OF
CC THE BLASTOCYST TO UTERINE EPITHELIAL CELLS AT THE TIME OF THE
CC EMBRYO IMPLANTATION. DIRECTLY RESPONSIBLE FOR HOMOPHILIC CELL
CC ADHESION.
CC -!- SUBUNIT: DIRECTLY BINDS BYSTIN, AND INDIRECTLY TASTIN.
CC -!- TISSUE SPECIFICITY: STRONG EXPRESSION AT IMPLANTATION SITES. FOUND
CC IN THE PLACENTA FROM THE SIXTH WEEK OF PREGNANCY. WAS LOCALIZED IN
CC THE CYTOSOL OF THE SYNCYOTIOTROPHOBLAST IN THE CHORIONIC VILLI
CC AND IN ENDOMETRIAL DECIDUAL CELLS AT THE UTEROPLACENTAL INTERFACE.
CC AFTER WEEK 10, THE LEVEL DECREASED AND THEN DISAPPEARED FROM
CC PLACENTAL VILLI. ALSO FOUND IN MACROPHAGES.
CC -!- SIMILARITY: CONTAINS 1 MAGE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB029037; BA83066.1; -
CC EMBL; AL049732; CAB86651.1; -
CC EMBL; U04811; AAA79334.1; -
CC Genbank; HGNC:12326; TRO.
CC MIM: 300132; -

```

DR InterPro: IPR002190; MAGE.
DR Pfam: PF01454; MAGE; 1.
DR PROSITE: PS50838; MAGE; 1.
KW Cell adhesion; Antigen; Repeat.
FT DOMAIN 400 598 MAGE.
FT DOMAIN 707 1381 69 x 10 AA APPROXIMATE TANDEM REPEATS.
FT CONFLICT 653 653 A -> S (IN REF. 3).
FT CONFLICT 692 692 G -> R (IN REF. 3).
FT CONFLICT 694 694 S -> G (IN REF. 1).
FT CONFLICT 724 724 S -> G (IN REF. 1).
FT CONFLICT 768 768 S -> G (IN REF. 1).
FT CONFLICT 780 780 S -> G (IN REF. 1).
FT CONFLICT 982 982 N -> G (IN REF. 1).
FT CONFLICT 1012 1012 S -> G (IN REF. 1).
SQ SEQUENCE 1387 AA: 138804 MW: 412015754B5A416 CRC64:

Query Match 4.7% Score 203.5; DB 1; Length 1387;
Best Local Similarity 22.6%; Pred. No. 0.00067;
Matches 192; Conservative 96; Mismatches 341; Indels 221; Gaps 39;

QY 53 GIADNA-----GWHNMGIDAVALDP-ODDQKYAAVGM-----YTNSMSPSNGA 95
DB 624 GIGEEAVAGRWNNDDMDICLTREELGDQAQMSRFSFEIARAQENADASTNVNFSRGA 683
QY 96 IIRS--SDRGATWSFTNLPKVGGMPCRGAGERLAVDPANS-----NIIFYGARGSN 146
DB 684 STRAGFSD-CASISIFNGAPSSSCGF--SGGPGITFGVAPSTASFSNTASISFGCTLSTS 740
QY 147 GLWKST-----DGVYFSKVSFETATGTYIIPDSDSNGYNSDKOGILMWTFPDTSS 197
DB 741 SSFSSMAASISFGCAHSTSTFSSEASISFGGMCTSNASFGVSSFSFGLP-----STSA 795
QY 198 TTGGATSRIFVGTADNITASVYVSTNASTWSAVPGOPGKYPFHAKAKLOPAEKALYLYTS 257
DB 796 TFGGASGSGFGR-----LSTTAG-FSGV-----LSTSTS 824
QY 258 DGTGPTDGLGSWRKDIAGTWTKDTTPVSGSLYFPGGLGLDLOKPGTLVYASLSNMW 317
DB 825 FGSAPTSTVFS-----SALSTSTGFGG-----ILSTSVCF 855
QY 318 PDAQLFRSTDSGTWTWSPINAMASYPET--YYVISTPKAPWIKNFIIDVTSSEPSDGLIK 376
DB 856 GGSPPSSSGSPGGLSTSLCEGSGPCTSTGFGGLST-----SVFSGSSSTSANFG--- 906
QY 377 RLGMWLESLEIDPTDSNMHLYGTGMTIFGSHDLTJNMDTRHNVSIOSLADGIEEFSYQDLA 436
DB 907 --GTLSSTICFDGSPST-----GAG--FCG-----ALNTSASFGSVLNTS 942
QY 437 SAPGSELLAAVDDNGFTFASRNDLGTSPQYVWATPTWATSTVDYAGNSVKSVYRVGN 496
DB 943 TGFSGAMSTSA--DFGGTLSTSVCFGSGPCTSVSFGS--ALNTNAGY--GGAVSTNTDFGG 997
QY 497 TAGTOVAISSDGGATWSIDYADTSMN--GGTVAYSAD--GDTILMSTASSG--VQRSQFOG 552
DB 998 TLSTSVCFGSGPSTSAFGGALNTNASFCAVSTASFGCAVSTSAFCGAPITNPFCG 1057
QY 553 SPASVSLPAGAVIASD---KKTNSVYVAGSGST--FYVSKDTGSSFTRGPK---LGS 602
DB 1058 AFSTASGFGGALSTADFGGTGPSNLSICFGAAPSTSVFSGAHGTSILCFGAPSTSLCFGS 1117
QY 603 ACTIRIDIAHPPTAGTLVYSTDVGIF--RSTDSGTFPGQVSTALNTYQIALGVSGSN 659
DB 1118 ASNTNLCPFGPSTSAFCGATSPSFCDEPSTSTGFSFG-----NGLSTNAGFGGLN 1170
QY 660 WNLVAFGTGPGSARLYASGDSGASWTD--IQSGGFGSIDSTKV---AGSGTAGOVYWG 714
DB 1171 TSA-GRGGGLGTSAGSGGLSTSSGPDGGLGTSAGGCGPGTSTGFGGLGTSAG--FSG 1227
QY 715 TNGRGVYAQGV-----GGGTG-----GTS----- 735
DB 1228 GLGTSAGFGGGLVTSDFGGGLCTNASFGLTGLTSAGFGGLSTSDGFGSRPNASFDRL 1287
QY 736 -----SSTKSSSSTSSASSSTTLRSSVSTTRASTVTSRTSSAAGPTGCVAGHYA 788

DB 1288 STIICFGSGSMTSTGFTGEPSTSTGFSGSPSIVGFGSPSTGVGFCGSPSTGFSGSPS 1347
QY 789 QCGIGMTGP 798
DB 1348 TGAGFG-GGP 1356

Search completed: April 26, 2003, 13:36:08
Job time : 38.2736 secs

Db 84 HDKIDALATDPVDTDRVYVAVGMVTNENDPNVGSILRSTQDGTWTERKLPKVGGMNP 143
OY 121 GRGAGERLAVDPANSNIIYFGARSGNGLKSTDGVTFSKVSFFATGYIIPDPSNGY 180
Db 144 GRMGRLAVDPKNSILIFGARSGHGLKSTGYATWSVTFWTGTYFQDSSST-Y 201
OY 181 NSDKGLMVTEDSTSTGGATSRIFVGTADNITASVYVSTNAGSTWASVPGPG-KYF 239
Db 202 TSDPVCIAWTFEDSTSGSSGATPRIFGVAD-AGKSVFSESDAGATMAVSGEPQYGL 260
OY 240 PHAKLOPAEKALYLYSDGTGYSVMYDIAGVWKITVYSGDLTFGFCGLG 239
Db 261 PHKGVLSPEKTLTISYANAGAPYDGTNHRKNTSGVMDISPTSLATYVGGGLS 320
OY 300 LDLOKPTLVASLNSMWPDAQLFRSTDSGTWSPIMAMASYPTERYYSISTPKAPMK 359
Db 321 VDLQVGTGLMVAALNCPMDELFRSTDSGATSPIMENKGYISINYYSYDLSNAPWIG 380
OY 360 NNEIDVTSSEPSDGLIKRLGKMIESLEIDPTDSNHLVGTGMTIFGCHDLTNMDTRHNS 419
Db 381 ---DPTS---TDQFPVRGMVMEALIDPFDGNHMLYGTGLVYGGHDLTNMDSKHNYT 433
OY 420 IOSLAGEIEFSVQDLASAPGSELLAAGDDGFTFASNDLGTSPOYMAATPYATST 479
Db 434 VASLAVGIEEMAVLGLITPPGCPALLSAGDDGFT---YHSDDAAPNOAYHTPTGYTN 490
OY 480 SYDVAGNSVKSIVRVGNTAG-TQVAISDGGATSIDYAADTSMNGCTVAYASADGDTILW 538
Db 491 GIDYAGNKSNTIRSGASDDYPTLALSSNGSTWYADYASTGTGVALSADGDTVLL 550
OY 539 STASSGVORSOGFASVSSLPAGAVIASDKKTSNVFAGSGSTFYVSKDGSFTKCP 598
Db 551 MSSTSCALVSKSGTTLAVSSLPAGAVIASDKKNTVFYFGSGAGIYVSKNTATSTFTKYV 610
OY 599 KIGSACTIIDIAHPPTAGTLYSTDVGIFRSTDSGTGQVSTALNTYQIALVGS-- 656
Db 611 SLSSSTTVAIRHPSIAGVMASTKGLHSHSTDYSTFQISGVTAGSFEFGKASST 670
OY 657 GSNMNLVAGTGPFGARLYASGDSGASMTDI-OGSOGFSGIDSTKYAGSGSTAGQVYGT 715
Db 671 GSVVLYGFTTIDGAGLFRSESDAGTNMOYISDASHGFGSGANVYNGDLOTGRVFRCH 730
OY 716 NGGVYTAQ-----GTVGCGTGT-----SSSTKOSSSTSSASSSTTLR---SS 757
Db 731 ERPGHLLRSORPARHBDGDDDTTSTKSTTVSTTLKTTSSASTSSSTVKTTS 790
OY 758 VVSTTRASTVTSRTSSAAGPTGSG--VAGHYAOCGGIGMTGPTQCVAPYVQKNDY 814
Db 791 SSTTSSASTTTTKTTTSTTTSTSSGTTATASAYAOCCGNGMTGATVCFGTCTTYSNAT 850
OY 815 YOCV 818
Db 851 SOCV 854

RESULT 2
O9P4T8 PRELIMINARY: PRT: 806 AA.
AC O9P4T8; 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE CEL6 protein precursor.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OX Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN SEQUENCE FROM N.A.
RC STRAIN=D649;
RA Morales-Almora P., Thurston C.F.;
RT "Molecular analysis of the cellulolytic genes in Agaricus bisporus.";

RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ292929; CAC02964.1; -
DR HSSP: P00725; 2CBH.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR002860; GH_BNR.
DR Pfam: PF02012; BNR; 8.
DR Pfam: PF00734; CBM_1; 1.
DR SMART: SM00236; FCBD; 1.
DR PROSITE: PS00562; CBD_FUNGAL; 1.
KW Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 806 CEL6.
SQ SEQUENCE 806 AA: 84779 MW: 76678326262PF341 CRC64;
Query Match 51.5%; Score 2235; DB 3; Length 806;
Best local similarity 51.7%; Pred No. 4.9e-91;
Matches 426; Conservative 126; Mismatches 220; Indels 53; Gaps 13;
OY 1 AFSMKNVKLGGGGFVPGIIFHRKTVAVARTDGLRLNADSWTAVTGDIADNAGM 60-
Db 27 SYRWQVKKIGGGGFVPGIIFHRKTVAVARTDGLRLNADSWTAVTGDIADNAGM 85
OY 61 HNMGIDAVLADPDQDKVYAAVGMVTNSWDPSNGALIRSDGATWSFTNLEPKVGGMNP 120
Db 86 NMGVYDALADPDQDKVYAAVGMVTNSWDPSNGALIRSDGATWSFTNLEPKVGGMNP 145
OY 121 GRGAGERLAVDPANSNIIYFGARSGNGLKSTDGVTFSKVSFFATGYIIPDPSNGY 180
Db 146 GRMGRLAVDPKNSILIFGARSGHGLKSTGYATWSVTFWTGTYFQDSSST-Y 205
OY 181 NSDKGLMVTEDSTSTGGATSRIFVGTADNITASVYVSTNAGSTWASVPGPGKYP 240
Db 206 NSDKIGIAWTFIDKAGSSGATPRIFGVANAGNSIYSTNGSSWAGVAGQPSFLP 265
OY 241 HRAKLOPAEKALYLYSDGTGYSVMYDIAGVWKITVYSGDLTFGFCGLG 300
Db 266 HNGVLSPEKTLTISYANAGAPYDGTNHRKNTSGVMDISPTSLATYVGGGLS 320
OY 301 LDLOKPTLVASLNSMWPDAQLFRSTDSGTWSPIMAMASYPTERYYSISTPKAPMK 360
Db 326 DQKNTVVAALNCPMDELFRSTDSGATSPIMENKGYISINYYSYDLSNAPWIG 385
OY 361 NNEIDVTSSEPSDGLIKRLGKMIESLEIDPTDSNHLVGTGMTIFGCHDLTNMDTRHNS 420
Db 386 NIDVAT---PGN---LOIMMESLISIDPDSNHLVGTGTYGSRDLKMDSAHNVTI 439
OY 421 IOSLAGEIEFSVQDLASAPGSELLAAGDDGFTFASNDLGTSPOYMAATPYATST 480
Db 440 KSLADGVEETSVQALISPSGPELVSAIDVGGEVHQS---LTRAPSAOFTPKMSTAD 496
OY 481 VDVAGNSVKSIVRVG---NTAGTQVAISSDGGATWSIDYAADTSMNGCTVAYASADGDTIL 537
Db 497 IDPAGNNPNVVRIGTGDSTSGKQVAISSDYVTMNHGAGDNDVGGVALSADNDIIL 556
OY 538 WSTASSGVORSOGFASVSSLPAGAVIASDKKTSNVFAGSGSTFYVSKDGSFTKCP 597
Db 557 WRTNGGVYVSRNOATF-----NIFYGASGKFTFYSTNGKTFSSAH 597
OY 598 PKGSAGTIRDAIHPPTAGTLYSTDVGIFRSTDSGTGQVSTALNTYQIALVGS 657
Db 598 GSLGATSVFDITVHPSVSGDIWASTDKLFHSTDSGATFSATIS-GITAMGVADAPPS 656
OY 658 SNMNLVAGTGPFGARL-YASGDSGASMTDI-OGSOGFSGIDSTKYAGSGSTAGQVYGT 714
Db 657 TGGYPAVFAANVAGGEVAFRRDDRGVNMVKLINDAHGGAASANMADPRVYGRVYIG 716
OY 715 TNGRGVYTAQGTGCTGCTSSSTKOSSSTSSASSSTTLRSSVSTTRASTVTSRIS 774
Db 717 TNGRGIFY--GDVAG-----SAPPTSATSTTSVST-----TTRSTTTTSTTT 761
OY 775 AAGPTGSGVAGHYAOCGGIGMTGPTQCVAPYVQKNDYVYOCV 818
Db 762 TSSGNGSPVSG-FCQCGGGGWTGTTCTVNGFQCVANPNYSOCI 804

RESULT 3

08TFPI PRELIMINARY: PRT: 857 AA.
 AC 08TFPI: 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Endoglucanase C.
 GN EGIC
 OS *Aspergillus niger*.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; *Aspergillus*.
 OX NCBI_TaxId=5061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21914059; PubMed=11916668;
 RA Hasper A.A., Dekkers E., van Mil M., van de Vondervoort P.J.I.,
 RA de Graaf L.H.;
 RT "Egic, a New Endoglucanase from *Aspergillus niger* with Major Activity
 RT towards Xyloglucan";
 RL Appl. Environ. Microbiol. 68:1556-1560(2002).
 DR EMBL, AY040839; AK77227.1; --
 SQ SEQUENCE 857 AA; 90459 MW; 482253ECD919C511 CRC64;

Query Match

Best Local Similarity 50.9%; Score 2211; DB 3; Length 857;
 Matches 432; Conservative 133; Mismatches 237; Indels 36; Gaps 14;

QY 1 AFSWKNVLGGGGGVEYPIIFPKTKVAVYARTDIGLYRLNADSWTAVDGIADNAGM 60
 DB 24 AYTWKVNVVGGGGGFTPGIVFNPASAKVAYARTDIGAVRLNSDDTWTPLMDW-ANNSMW 82
 QY 61 HMKGDVAVALDDDDQKVAAYGMYNMSWDPNSGATIRSDGATSFNLPKVGGNP 120
 DB 83 HMDGIDAIATDPDDVDVVAAGMYNDPNDGSLRSTGDTWEEKLPKVGGNP 142
 QY 121 GKGGERLAVDPANSNIIFGARGSGNLKSTGVTGFSSTFATGTIIPDSNGY 180
 DB 143 GKGGERLAVDPNDNSILFGARGSGNLKSTGVTGFSSTFATGTIIPDSNGY 200
 QY 181 NSDKGGLMVTVDSTSTTGATSRIFVGTADNITASVYVSTNAGSTWSAVPGP-GKXF 239
 DB 201 TSDPVGIAMVTFDSTSGSGSGSPTRIFVGVDT-GESEVSEADGETMWSGEPYGL 259
 QY 240 PHKAKQPEKALYLYTSGTGYDSTGLSVKRYDINAGTKMDITVSSDLYFGGGG 299
 DB 260 PHKGLSPSEHLYTSGAGAGYDGTNGTVHRYNTTSGVWTDISPTSMITDYYGGGLA 319
 QY 300 LPLQKPGTLVAVSLNSMWPDAOLFSTDSGTWSPIMANASPTETYYYSISPKAPWIK 359
 DB 320 VDLQVGTMAVALNKMWDDELIMRSTDSGCTSPIMANNGPSIYYISYDINAPWIK 379
 QY 360 NNEIDVTSPSDGLIKRLGMIESLEIDPTDSNHLVYGTGMTIFGGHDLTMDTRHNS 419
 DB 380 D-----DTSTDEFEVRYGMWEALAIIDPDSDHLYGGETIYGGHDLQNMDSHNTV 432
 QY 420 IQLADGIEFEVQDLASAPGSELLAAGDNGFTFPAKNDLGTSPQVMAPTATST 479
 DB 433 ISLAVGIEEMAVLGLTPPGALLISAVGDGCFHTS--LTPASOYVYHPTYSSTN 489
 QY 480 SVYDAGNSVSVVAVGNT-AGTQVAISDGGATWSIDYADOTSMMNGTAVASADGTILM 538
 DB 490 GIDYAGNKRANIRVSSSDSDPTLALSSFGESWYADYAASSSTATGOVALSDADTITLL 549
 QY 539 STAAAGVQNSQFGSFASVSLPAGAVIASDKTKNSVYFAGSGSTFYVSKDGSSTFGP 598
 DB 550 MN-SDAVRSANSATLSAVSLPSGAVIASDKANNNTFYFAGSAGSSTFYSSDVAATFTVT 608
 QY 599 KIGSAGTIIDIAHPTTACTIYVSTDVGFIRSTDSTRGQVSTALNTYQIALGVS-- 656
 DB 609 TIGSSTTANAIKOPSLADWVSTDTGLFHSITNGKSTFTQIGSGCTEGWSFGFGKPSSD 668

QY 657 GSNMNLVARGTGPBGARLYASGDSGASMTDIOGSG-FGGSIDSTKVGASGTAGQVYCT 715
 DB 669 GDYPLVLFGEFTYDVGTLGFKETDQGVNMQIISDAEGRFSASANYVNGLOVGRVFT 728
 QY 716 NGRGVFYA--QGTVGGT-----GGTSSST-KOSSSTSSASSSTTL-RSSVY 759
 DB 729 NGRGIFYCDPSGTLPLSTAVSSASTAVKSSSTSTSKVSGSTVYSSSTATTTTSSIK 788
 QY 760 STTRASTVTSSTSSAGPTGSGVAGHYAQCIGIGTGTQCVAPVYCKODYVYQC 817
 DB 789 STTLTNTKSSSTSTSTSTATGTASAVGCGCGSGTGTGCPGSCWTCYENEVYSQC 846

RESULT 4

086727 PRELIMINARY: PRT: 890 AA.
 AC 086727: 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Putative secreted cellulase.
 GN SC06545 OR SC5C7.30C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxId=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Parhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinovich S., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RT Nature 417:141-147(2002).
 DR EMBL: AL031515; CAA20642.1; --
 DR HSSP: P07986; 1EXG.
 DR InterPro: IPR001919; Bac_cellose-bind.
 DR InterPro: IPR002860; GH_BNR.
 DR Pfam: PF02012; BNR_10.
 DR Pfam: PF00553; CBM_2_1.
 SQ SEQUENCE 890 AA; 93252 MW; D2C58695A4B56E84 CRC64;

Query Match

Best Local Similarity 36.6%; Score 1591.5; DB 16; Length 890;
 Matches 345; Conservative 134; Mismatches 286; Indels 99; Gaps 23;

Dh 680 DASD---TVGLCKSKTDDGCPAIIYMDATIDGTAGIFRSDDEGATWIRINDAHQYGSPPY 736
Qy 699 TKVAGSGSTAGOVYVGTNGRGVFAVAGTVGG 729
Dh 737 C-ITGDPNKKYGRVFCVGTNGRGVIV--GDIDG 764

RESULT 6

Q9AOH0 PRELIMINARY; PRT; 996 AA.
AC Q9AOH0:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Glycosyl hydrolase 5 (Fragment).
OS Caldicellulosiraptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Syntrophomonadaceae; Caldicellulosiraptor.
OX NCBI_Taxid=80339.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-TOK7.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
Bergquist P.L.,
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiraptor isolate Tok7B.1."
RL Curr. Microbiol. 40:333-340(2000).
DR EMBL: AF078038; AAK06388.1;
DR HSSP; Q06851; INBC.
DR InterPro: IPR001956; CBD_3.
DR InterPro: IPR002860; GH_BNR.
DR InterPro: IPR002965; P_rich_extensn.
DR Pfam: PF02012; BNR. 9.
DR Pfam: PF00942; CBM_3; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR PRODOM: PD001947; CBD_3; 1.
KW Hydrolyase.
FT NON_TER
SQ SEQUENCE 996 AA; 108275 MW; 3C72B6ED22F3C614 CRC64;

Query Match 33.5%; Score 1454.5; DB 2; Length 996;
Best Local Similarity 38.2%; Pred. No. 1.5e-56;
Matches 306; Conservative 142; Mismatches 305; Indels 48; Gaps 19;

Qy 2 FSWKVKVLGGGGFVPGIIFHPKTKGAVYARTDGLGLR-LNADDSWAVYDGIADNMGW 60
Dh 37 YWKNVWKTIEGGGGFTTGLVFNPKENLVYRTDGGATRTDGGNTWQLMDWVSFD-DW 95
Qy 61 HNWGIDAVLDPDODOKY-YAAGVWYTNMSPNSNGAIIRSSDRGATWSFTNLPKVGGM 119
Dh 96 NLUGVESTATDPVDNXYIILACOGTYTNSWMDMNCALIRSDGTFELTLPFLGLGGM 155
Qy 120 PERGAGERLAVDPANSNIITFGARSGNGLKMSDGVTSFVSKYSFTATGYIIPDSDSNG 179
Dh 156 PXRNIGERLADIPNNRILYLGTREGXGLMKSEGYGVWKKVTSFPNPGTYIEDNCPND 215
Qy 180 YNSDKGGLMWTPEFDSSTSTGGATSRIFVGTADNTASVYVYTNMGSTWNSANPGQPKYF 239
Dh 216 YLNHTTGVVWVFPDTSRPGEGKIIITVGVADK-TTISITTKKDGQWQALPGQPTGLL 274
Qy 240 PHKAKLPAERKALVLTYSDGTEGPTDGLGVSVMRYDIAGGTWMDIPVSGSDLYFSGGIG 299
Dh 275 PQRAKLS-SDGLVITYITNTGSPYNGDYGWVRVYTKTGEMKINISPMMAQDITFYGGGLA 333
Qy 300 LDLOGPGLVAVASLNSMPPDOLFRSDSGTWSPIWAMASYPTETYYISSTPKAPWIK 359
Dh 334 YDAQNPKVYVMAALSSMPDITVIMRSIDGGETWKCIMWENGYPNRLTYLNMDSAAPML- 392
Qy 360 NNEIDVTSSEPSDGLIKRLGWMIESLEIDPNDNSMHLVGTGTFITGGHDLTWMDRHNVS 419
Dh 393 -NF-GTBPPTPE-VSPKLGKRWVGTLEIDPFNSDXMLGTGAXLYGCDLITWMDKGNIT 449

Qy 420 IQSLADGIEEFSVODLASAPGSELLAAVGDNGFTFASRNDLGTSPTQVATPTMATST 479
Dh 450 IKVKAIGIEEFSVQALLISPPVGPLHLSALGDIAG---RHEDLEKAPWMTYVQPMGTTT 506
Qy 480 SVDYAGNSVKSRYRGNT-----AGTOVAISSDGGATWSIDYADT-----SMNGYAY 529
Dh 507 DIDFAELNPNFMVRYGVNDKQMPNPTNRIGFSYDQKSM---FOGTEPOSTSEGVVAA 563
Qy 530 SADGDTIIMSTASSGVQSRQFQGS-FASVSSLPAGAVIASDKKTNSVYAGSGSFYVSK 588
Dh 564 AADGSAVVAWAPKGAIVCYSTDNGKMKVCEANPSEALIVSDRVNPKYARKNGKFTISA 623
Qy 589 DTGSEFTGPKLG--SAGTIRIDIAHPHTAGTLVSTVGVIFRSDSGTTEGCVSTALTUN 646
Dh 624 DKGRFIESPAAGLPIISGNFTV---PEIEDIMLVGNNGMHISDGGYSVKISGV--- 677
Qy 647 TYQIALGVSGSSNNN-----LYAFGTGSGARLIVASGDSGASWTDIQSGQSGIDS 698
Dh 678 --EDAASTGFGRPARGETYPAIYTYAKINGVRC--IFRSDCDKTKWIRINDKKHOFCCAN 733
Qy 699 TKVAGSGSTAGOVYVGTNGRGVFAVAGTVGGTGTGTSSTQSSSTSSASTLRSSV 758
Dh 734 ADITDPPRYGKRVFAVATGLIKW--GEIATVSNISPSATPTSTPFTPTPTATITSTP 791
Qy 759 VSTRASTVTSKRTSSAAGPT 779
Dh 792 APTSTPTPTPTPTPTPTPTPTPT 812

RESULT 7

Q9WYE1 PRELIMINARY; PRT; 707 AA.
ID Q9WYE1:
AC Q9WYE1:
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Endoglucanase, putative.
GN TM0305.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_Taxid=2336;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima."
RL Nature 393:323-329(1999).
DR EMBL: AE001712; AAD35393.1;
DR TIGR: TM0305;
DR InterPro: IPR002860; GH_BNR.
DR Pfam: PF02012; BNR. 9.
KW Complete proteome.
SQ SEQUENCE 707 AA; 79496 MW; 8CD8743CBDA6A99F CRC64;

Query Match 20.7%; Score 899; DB 16; Length 707;
Best Local Similarity 32.7%; Pred. No. 2.9e-32;
Matches 248; Conservative 118; Mismatches 282; Indels 110; Gaps 33;

Qy 2 FSWKVKVLGGGGFVPGIIFHPKTKGAVYARTDGLGLRLNADDS-WTAVYDGI-ADNAG 59
Dh 21 FEWKSVEI-NGGGEVPGIIFHPASGLIYARTDVGGLRMBEETKRWKQLDFLRDQSD 79
Qy 60 WNWGIDAVLDPDODOKY-YAAGVWYTNMSPNSNGAIIRSSDRGATWSFTNLP---FVYG 116
Dh 80 Y--MGVLSVALDPSDPKRIYAWTKGYTDW--AGYCALLISEDYGETWTVINDKGYKIVG 136

Oy	117	GNNPGRGCEPLAADPANNSNITVFQARSGCKLWKSRTDGCVTSSKSPFATGTIIPDS	176
Db	137	GNEGGRRANGERLOYDPAPFSSVLFNGCT-TKGLKSKSDFGKNMKKDKDSPST-----	186
Oy	177	SNGYNSDKOGIMWTFEDSTSTTGATSRIFGVGTADNITASVYSTNAGSTSAVPGPG	236
Db	187	-----SVTFELPFDEKGEGEKSPTRIFVGCSE--PKGIFTVEDGTTNVNLPJLN	235
Oy	237	KYFHAKKLQPAEALKYLTYSDDTGPYDLGSVMRYODIAGCTHWDTTPVSSSXLFPFG	296
Db	236	DLPLRGRIH-DeILVTLISNALGPRGARGVMKTYVLADOKMVDVPKMGD----FG	288
Oy	297	GLGLDLKPGLTVLASINSMMPPDAQLFRSTDGTSPTSPIMAWASPYETERYYSISTPAP	356
Db	289	YCIGIDVOE-NVVIYSTLDKMYPHDEIFISLNGSETMRPLLEKANPDIN-----KAP	338
Oy	357	WIKNNFDIVTSESPPDGLIRLCGMIESELDIPTDSNMHLXGTGMTIFEGHDLTMMDRH	416
Db	339	WIK-----DLNPH-----W-ISDVLDIDPDMNRKAIFTTGYGVVVYTELRK----	377
Oy	417	NVSIOSLAD-----GIEEFSYODLASAPCGSELLAAVGDDGDFEPASRNDGTSQP	467
Db	378	--SEGNGKRPVKATFEKRGILEETVYLQVLPPIERRLSLADIWGCFRIES---LDTPPS	432
Oy	468	TWMTPTMATSTSYDVANGSVKSVYRVGNATGOVAISSDGATW-SIDYAADTMNG-525	
Db	433	SMYPPLMW-TSLGIAFYQNRSKFVARHTYTYPFLSYSEDGGINMEIETVEGITDGR	491
Oy	526	-TVVYASADGPIILMTSSGVSORSOPQG-SFAVSLSLA---GAVIASKTKNSVFA--578	
Db	492	LSLVANSDBGTLVWSPANHEIYVSSDKGSKMAAISVPPEENYFPASPVPVPSKFYFD	551
Oy	579	SGSGTFVYVKDTSSTGRCPKLS-----AGTRIADAHPPTAGTLVYSTD-VGIFRSTD	632
Db	552	WKNDPLISKDGKGSFKGAKLPDSFMWVNLVSPVLPADRDGDIMIALONNGLYRMD	611
Oy	633	SGTF---GOVSTALTNTYQIALGVGS-GSNW-NLAFGTGSGARLVLAGSGSGASMTDI	687
Db	612	GGTFERLGANDVIA---YYIGCAPKPGTDYPAIYLNGVMGVYGIFMSSTDGKTWMI	667
Oy	688	OGSO-GFGSIDSTVKAGSGTAGOVYGTNGRCVFFAQ	724
Db	668	NNDKHFGWMH-YMIGDNMFGRIFLGTGCRRIIYGE	703
<hr/>			
RESULT 8			
P91365		PRELIMINARY;	PRT; 2232 AA.
AC	P91365;		
DT	01-MAY-1997 (TrEMBLrel. 03, Created)		
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)		
DE	K06A9.1 protein.		
GN	K06A9.1.		
OS	Caeenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabdilitida; Rhabditoidea;		
CC	Rhabdilitidae; Paludoderinae; Caeenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RA	Giesel C., Gating S.;		
RL	Submitted (JAN-1997) to the EMBL/Genbank/DDAI databases.		
CC	-1- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY		
CC	ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT		
CC	OF THE A FORM.		
DR	EMBL; U08046; AAC70889.1; -;		
DR	EMBL; U08046; AAC70890.1; -;		
KW	Alternative splicing.		
FT	VASSPLIC 842 866		
FT	VASSPAPTSQNPNPSTSSGSSMI -> LATSPAPRSVY		
FT	CLPMD (IN ISOFORM B).		
FT	PYPSOSTSVESSTTPSPGSPGTLTSTSPSPSOTTGIST		
FT	OGSTSPGISTTEEMTSOQTGPSTGSTSVTOGTSVDSST		
FT	VANSPPLIC 870 1051		

[illegible]

OY 759 -VSTRAST-----VTSRTSSAGPTGCVAG 785
 DB 1882 GVTSASSTQPMSTQSSAGSTVASTAG 1913

RESULT 9

O76602 PRELIMINARY; PRT; 1275 AA.
 AC 076602;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Hypothetical 122.9 kDa protein.
 GN H02F09.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodertinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 Garner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Shownkeen R.,
 Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 Thierly-Mieg J., Thomas K., Vaulin M., Vaughan K., Waterston R.,
 Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Geisel C., Harmon G.;
 RT "The sequence of C. elegans cosmid H02F09."
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF077538; AAC64622.1;
 KW Hypothetical protein.
 SO SEQUENCE 1275 AA; 122924 MW; CB59B70C05959E25 CRC64;

Query Match 6.9%; Score 300.5; DB 5; Length 1275;
 Best Local Similarity 24.1%; Pred. No. 1.3e-05;
 Matches 172; Conservative 94; Mismatches 303; Indels 145; Gaps 26;

OY 162 SSFTATGYIPPPSDSNGYNSDKOGLMWTFPDSTSTGTCATRLFTVADNITASVYS 221
 DB 446 TSIGSSSTPLPSQSTSLMS-----LSTYTPSSSTAGATS---PAQGSTKPTIGIS 495
 OY 222 TNAGSTWMAVPOGPKYFPKAKLO--PAEKALYI-----TYSDCGTGYDGLGSVMRY 273
 DB 496 MSSGPT-TVAPGAS-----TESTVLOSSTPSPGTTVLLPSSGSTATAGTSPQASTVTVV-- 547
 OY 274 DIAGGKWKDITVPSGSDLYFGFGGLGLDQKPGTLVVAFLNMMWPDQQLFRSTDSGTTWS 333
 DB 548 -----TDISTVSGSTVSTQTAESLSTESP-----TSAGSSIS 580
 OY 334 PTWASVPTETYYYSISPKAPWIKNNFIDVTSSEPSDGLIKRLGMWTESIEDPTDSN 393
 DB 581 TVSTVSSQPS-ty-----IPVSAASSISYTLSTGSGTSPASGCTESSG-----SSTSGP 628
 OY 394 HMLYGTGMIIFGCHDLTNDTRHNVSIGSLADGIEEFSVQDILASAPGSGELLAAVGDDNG 453

DB 629 STISSASASTVGSIVTEASTISGTESTSTPGTESTVSE-ASTVSGSSVSTVSGSTES 687
 OY 454 FTFASRNDIGTSPQIVMATPWTATSTVDYAGNSVYKVRVGNATGATVAAISDDGATWS 513
 DB 688 TSAGASTVSGSTGTVSDSSTISDSTGNTNAPGSTESTVYTGSSVSTVSGSTSTGPSTMS 747
 OY 514 IDYADTSMNGTVAVSADGDTILMSTASSGVOR-----SOFQGSFASVSLIP 561
 DB 748 AS-TGSTNTPGSTESTITFDGSTVSGSTGSTNNPGSTDSSTTGISTVSGSSLSTISGS 806
 OY 562 ACAVIT-ASDKNTSVFYAGSGST-----FVSKDTSSTTRPKLGSASTINDIAH 612
 DB 807 TSGTSGSSDMDTVSTGTSPPSTESTVSGASTMSPSTGSEVETSGSGSVSTVSGSTSS 866
 OY 613 PTT-----AGTLVSTPVGIFRSDSGT-----FGVSTALNTVYIALGVG----- 655
 DB 867 STTGQSTVSESSVSTVSSSTISQSTGTTTDESVTFGSGTATAGSSWMASTGTDTP 926
 OY 656 -----SGSNMNLVAFCTGPGSARLYASG-----DSGASWT 685
 DB 927 GSTESTITGTVTGSSTVSGSTGSTTGESITSESTMTVGVSTGTTIGESTVSGSTRS 986
 OY 686 DTQG-SQGRGSIIDTKVAGSGS---TAGQVYGTNGRCVFIAGTYGGTGSTSSSTKOS 741
 DB 987 TVTGESTVSGSTEST-VSGSTESTPTVPSTVSGSTGTV-TGESTVSGSTASTSGSTGS 1044
 OY 742 SS-----STSSASSTLLRSSVSTTRA-STVTSRTSSAGPTGSGVAGHYACG 791
 DB 1045 STEAGSTVSGSSASIVTSTSGSTSGESTVSGSTVSTVSGSTGTTTGESTVSG 1098

RESULT 10

O98MG7 PRELIMINARY; PRT; 3145 AA.
 AC O98MG7;
 DT 01-OCT-2001 (TReMBLrel. 18, Created)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical glycine-rich protein mlr0587.
 GN MLR0587.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti."
 RT DNA Res. 7:331-338(2000).
 DR EMBL; AP002995; BAB48146.1;
 DR Interpro: IPR000087; Collagen.38.
 DR Pfam: PF01391; Collagen.38.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 3145 AA; 271616 MW; F10C4A98282766FB CRC64;

Query Match 6.8%; Score 294; DB 16; Length 3145;
 Best Local Similarity 23.1%; Pred. No. 6.8e-05;
 Matches 228; Conservative 116; Mismatches 364; Indels 278; Gaps 52;

OY 10 GGGGFPVPIIHPKTKGYAVARTDIGLYR--LNADDSMTAVTDGIADNAGMHWGIDA 67
 DB 239 GGIGAYV-----GVLVARTMGVDTNFMVLANV-NMWSVTGTGI-----GADV 278
 OY 68 VALDPODD-QKYAAVAGMT-----NSWDPNSGAIIRSSDKGATWSFTNL-----P 112
 DB 279 LTISSPTROPPLSTAGTTTITSLNIGSNTLTLINGAAVILAATSGITWASPAITGTGT 338


```

OY 113 FRVGNMPCRG-----AGERLAVDPANSNIIYFGABSGNLMK---STDG 155
DB 339 IAAGNITGAGTVGIPISITTTGTTASGGLNTLGTVSNRTLAIANVAGSILKLDGTVSSG 398
OY 156 VTFKSVSTATATGTYITPPDSDSNGNSDKGLMWTTFDSTSTTGATSRIRIVGADNIT 215
DB 399 VFN-----GTTGTL--DLTNLAGFTGTISGL-----NIGASTTATNQIDFGNASTVS 445
OY 216 ASVYSTNAGSTMSAVPGOPKYPFHAKLOPAEKALYLYTSDGTGPYDGTGSLVWRYDI 275
DB 446 GGL-----NGSVLTVYVDGAGOSY-----ALTL-----CGAPAGTYVDL-KFPG 483
OY 276 AGTWMKI--TPVSGSDLYFGFGGL-----GLDLKP-----G 306
DB 484 ASGT--DVEFFSTPATNTFTMAOGGADAWGDTARMSSGGLPLQPAAGPULTFVSNHNAOG 541
OY 307 TLVAVSLNMMWDALFRSDSGTMSPTIMANASPTETYYYSISIPKAPMLKN---FI 363
DB 542 SL--SGNINSLVMPPLRLQRTGTG---SSIMTVSDTGSSANQFTTSLT---LINHQITV 593
OY 364 DVTSSEPSDGLKRLGMMIE-----SLEIDPTDS----- 392
DB 594 NGTANLTS--GTATNTMWTGTGCVAGAGLGNQYFENDGSLNVTGTTAALANGFSTATF 652
OY 393 --NHMLYGTG-MTIFGHDL---TNMDTRHNVSIQSLADG-----IEFVSVDLASA- 438
DB 653 TGNIVYTGSGYNNIYGNASLVEFGTGVGASQITNFVTDGNNTDGTVEISSQFVNAAY 712
OY 439 ----FGCSLLAAVG-----DNGC-----FTFASRDLGT----- 464
DB 713 AGFLPGDTLVLENLGATPLSEVYIDGNGQATVYIYSGANGSTGLLDETFLENGFTNGSD 772
OY 465 -----SPQTVAT-----PTMAT-----STSYDAGNSKSVVR 493
DB 773 FTFVSSANGGVYTCATGTSAGAHVPTPTGPTGATGTSATGATGATGATGATGATGATG 832
OY 494 VGNTAGTOY--AISSDG--GATWSIDYAADTSMNGSTVYASADGDTILWSTASSGVORSO 549
DB 833 TGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 890
OY 550 ----FGCSASVSLPAGANVIAVDKTNVFAAGSGSTVYVSKDQSSSTTRPKLGSACT 605
DB 891 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 949
OY 606 IRDIAHPTACTLYSTDVGIFRST--DSGTGFGVYSTALTYQIALCVGSGSNMNLVA 664
DB 950 TGDGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 997
OY 665 FG--TGPSGARLYASGDSGASW--TDIQSGOG---FGSIDSTKVAS--SGSTAGOVYVGTNG 717
DB 998 TGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1056
OY 718 -RGVFAAGTVG--GGTGTSSSTKSSSSSTSSASSSTTLRRSVSTTASVYTSRTSS 774
DB 1057 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1116
OY 775 AAGPTG--SGVAGHYAOCGIGWTGPT 799
DB 1117 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1142

RESULT 11
O98E20 PRELIMINARY: PRT: 3930 AA.
AC O98E20:
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical protein ml14444.
GN ML14444.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

```

```

OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFEF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idekawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT *Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003004; BAB51100.1;
DR InterPro: IPR001425; Bac_rhodopsin.
DR InterPro: IPR000734; Lipase.
DR InterPro: IPR003880; Peptide_atlase.
DR PROSITE: PS00327; BACTERIAL_OPSIN_RET; UNKNOWN_1.
DR PROSITE: PS00120; LIPASE_SER; UNKNOWN_1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 3930 AA; 380662 MW; 9ACD1ACA185BF712 CRC64;

Query Match 6.7%; Score 290.5; DB 16; Length 3930;
Best Local Similarity 21.5%; Pred. No. 0.00012;
Matches 191; Conservative 110; Mismatches 303; Indels 283; Gaps 38;

OY 6 NYKLGSGGFGFVGIIFHPTKGV-----AVARD-----IGLYRLNAD 44
DB 747 NAEFSGGSGIGSTASLDAERLNOTLGHAFGSTTFSPOLFDQONTFVGVGSLGCG 806
OY 45 DSWTAVTDGIDNAGHNMGIDAVALDPDDQKYYAAGMYTNSMDSNGAIIRSDRGA 104
DB 807 SSTSSL--LAINLADANTQNNAIFFLVNEGSPIDA-----NGFSLSDGQTLFSGNCR 858
OY 105 TWSFTLPRKVGGM-----FQRCG-----GERLAVDPANSIIY----- 139
DB 859 SPSLGVLPINIGDVNVAIDPDAITLNSGGAVTYANGNSLIDFNISGSDAG 918
OY 140 FGARSGNGLMKSTDGVTFSKYSS---FTATGTYIPDPSDSNGVSKDGLMWTYFDS 195
DB 919 INATSTNGL---TIQGYTANVDITGLFSGVGTGYVD--NLNVQNASGTGLELGSMT 973
OY 196 SSTGGATSRIFVGTADNTTASVYSTNAGSTWASVAPGPKYFPKAKLOPAEKALYLT 255
DB 974 VNETGNAT--IGNATN-----GLSAN 993
OY 256 YSDGTPYDGTGSLVWRYDIAGCTWKDITPVSGSDLYFGFGGLGLDLQKPTGLVVASLNS 315
DB 994 NFDGTATFDD-----LDITGG-----GI--GVGYVNGSSGTLTFGA-- 1027
OY 316 WMPDAQLFRSTGSGTWTSPIMANASPTETYYYSISTPAPKIKNFIDVTSEPSDGLI 375
DB 1028 -----GSSIAIGTS-----SNAFSISNSTPNTV--NCTISQTAASAASKVS 1066
OY 376 KRLGMMIESLEIDPTDSNHMLYGTGWTIFGHDLTNMDTRHNVSIQSLADGIEEFSVODL 435
DB 1067 NMTG-----GTAFGGMITASTGAANLIDSTNAGCAINF----- 1102
OY 436 ASAPGSELLAAGDNGTFP-----SRNDLOT-----SPQTVWATP 473
DB 1103 ---GGDLTLA---DGFASVAGGGLTYVSGNTVATGLGPDEKGIQIGMTIGANG 1154
OY 474 TWATSTVDYAGNSKSVVRVGNCTAGTOY-----AISDGGATWSIDY----- 516
DB 1155 VNFSSVAVVDSGGGVSTGAIINSTSGANTFPGVDFRISGTAISLDNMAASGTYSNCTT 1214
OY 517 AADTSMNGG-----TVAYSADGDTILWSTASSGVORSOFGSFAVSLSLAPAGVIASD 569
DB 1215 KVDTVVGGPGLVQNSAATVNVGNLVTNINAGADVSLTNMTG-----TIGIGGTTNS 1268
OY 570 KTNISVFAVAGSGSTFVSDTGSSTFRGKLSAGTIRIDIAHPTAGTLVYSTDVGIFR 629

```


Db 2339 TARPSVAGSGTGTGVSAGAGSGTSSGSPGATCASIGQPETSRISVAGSGAPAV-SSGA 2397
 QY 773 SSAAGTSGSGVAGHYAAGC-----GIGMTG 797
 Db 2398 SOAAGTSGAGPCTTASVGTETARPVSAGSGTTG 2432

RESULT 15

Q9VTK8 PRELIMINARY: PRT: 2586 AA.
 AC Q9VTK8;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE CG18331 protein.
 CN CG18331.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed-10731132;
 RA Adams M.D., Celinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-fiankooch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktiroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova K.C., Botchan M.A., Bouck J., Brokstein P., Brothier P.,
 RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Goddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Houslin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relmet K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodcock T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003544; AAF50040.2;
 DR FlyBase: FBgn0036181; CG18331.
 SQ SEQUENCE 2586 AA; 260194 MW; 8EBB2435A9FEAE5B CRC64;

Db 173 OCTNGNSSTQSSSTTTTSSDEQOTSSDPVEVAGSSSNGDNGNSTQSLTTT 232
 QY 195 TSSITGAGATSRFLVAGADITASVYSTAG-----STWAVAGGQKTFPHNA 243
 Db 233 TTSSDGGQST-----TSSDPVEVSGGTNGNSSTQSSSTTTTSSDEQOT-----TS 281
 QY 244 KLOPAEKALYLTYSDDGTGPDYDGLGSMWRDYDAGTGWKIDTPVSGSDLYFGFGGLDLQ 303
 Db 282 SSDPVEVAGSSSNGD-----NSTQSS-----TTTTTTTTSSD-----GGOSTLS 325
 QY 304 KPGTLVVASLNSMWPDAOLFRTSDGTTWSPIMAMASYPLETYYSISIPKAPWIKNFI 363
 Db 326 DPVEVSGGTNGNSSTQSSSTTTTSS-----SDEQOTSSDPVG----- 367
 QY 364 DVTSESPDGLIKRLGKMLIESLEIDPTDSNMHLXGKMTIF-----GGHDLTNDTRIN 417
 Db 368 EVAAGSSSNG-----DGNSTQSSSTTTTSSDGGQSTSSDPVE 410
 QY 418 VSIQSLADGIEEFVSQDLASAPGSELLAAVGDNDGTFPAS-----RNDLQTSPO 467
 Db 411 VS-QGTNG-----NSTQSSATTTTSSDEQOTSSDPVEVAGSSSSTGDS 461
 QY 468 TWANPTWATSTSVYAGNSVKS--VVRV-----GN-----TACTQVAISSDGATWS 513
 Db 462 TOSTTTTSTTTTSSDGGQSTSSDPVEVSGGTNGNSSTQSSSTTTTSSDEQOTSS 521
 QY 514 -----IDYADPMSMGCTVAYASADGTLIMSTASSSQVQSGQASVSSLPACAVITS 568
 Db 522 SSDPVEVAGSSSNG-----DGNSTQSSSTTTTSSDGGQSTSSDP--VVEA 571
 QY 569 DKTNVVFYAGSGST-FVYSKDTGSFTRPKGSAGTIRIDIAHPTTAGTLVYSTDGI 627
 Db 572 SGGT-----GMSSTQSSSTTTTSSDGGQSTSSDPVEVAG-----SSSID 618
 QY 628 FSTDSGTTFGGVSTALINTYQIALGVGSGSMNMLYAGTGPBARLYVAGSGASMTD 686
 Db 619 GNSTQSTT-----TTTTTT-----SSDGGQSTSS 645
 QY 687 ----IOSQGF-GSIDTKVAGSGSTA-----GQVYVGTNGRCGYFAQGVGGTG-GTSS 736
 Db 646 DPVEVSGGTNGNSSTQSSSTTTTSSDEQOT-TSSSDPVEVAGSSSNGDNGNSTQSS 704
 QY 737 STKQSSSTSSA--SSSTTLRSSVV-----STPRASTVTSRTSSAGPTGS 781
 Db 705 STTTTSTTTTSSDGGQSTTSSDPVEVSGGTNGNSSTQSSSTTTTSSDEPTTS 761

Search completed: April 26, 2003, 13:37:18
 Job time : 67.7061 secs

Query Match 6.0%; Score 260.5; DB 5; Length 2586;
 Best Local Similarity 24.1%; Pred. No. 0.0016;
 Matches 173; Conservative 78; Mismatches 253; Indels 213; Gaps 32;

QY 150 KSTDGCVTFSKVSVSTANGT-----YIPDPSDSNGYNSDKGLMWTFTDS 194

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:26:02 ; Search time 2.42891 Seconds
(without alignments)
1042.344 Million cell updates/sec

Title: US-10-026-994-3

Perfect score: 88
Sequence: 1 MKVSRVLAIVLCAVIPAAH 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	54.5	368	22	ABG13124
2	46	52.3	242	18	AAW10017
3	46	52.3	333	16	AAW72885
4	46	52.3	333	21	AAV90613
5	46	52.3	333	21	AAV90647
6	46	52.3	333	22	AAU01295
7	46	52.3	333	22	AAU01296
8	46	52.3	333	22	AAU01297
9	46	52.3	333	22	AAU01298
10	46	52.3	333	23	ABG5918

11	46	52.3	333	23	ABB4683	Human GPR8-11gand
12	46	52.3	333	23	ABB4723	Human GPR8-11gand
13	44	50.0	1819	22	ABG28734	Novel human diago
14	44	50.0	2267	22	ABG13742	Novel human diago
15	43.5	49.4	776	21	AAV78516	Novel human diago
16	43	48.9	77	23	ABP43537	Isoamylase amlno a
17	43	48.9	233	19	AAW40075	Guinea pig eosinop
18	43	48.9	430	23	ABB57124	Mouse ischaemic co
19	42	47.7	114	22	ABB43495	Peptide #11001 enco
20	42	47.7	114	22	ABB26455	Protein #8454 enco
21	42	47.7	114	22	AAW64429	Human brain expres
22	42	47.7	114	22	AAW77239	Human bone marrow
23	42	47.7	114	22	AAW21175	Peptide #7609 enco
24	42	47.7	114	22	AAW37388	Peptide #11425 enc
25	42	47.7	114	23	ABG46252	Human peptide enco
26	42	47.7	188	22	AAU01629	Human secreted pro
27	42	47.7	365	14	AAW44235	Human secreted pro
28	42	47.7	365	14	AAW38157	HTLE receptor. Sy
29	42	47.7	365	22	AAE00866	Sequence of the hu
30	42	47.7	483	11	AAW07445	Human 5-hydroxytry
31	41.5	47.2	160	22	ABB11168	Secretory signal p
32	41	46.6	183	22	AAU48211	Human secreted pro
33	41	46.6	214	21	AAV75578	Propionibacterium
34	41	46.6	214	21	AAV75579	Neisseria meningit
35	41	46.6	254	22	AAW81249	Neisseria meningit
36	41	46.6	346	18	AAW23088	Human AFP protein
37	41	46.6	365	22	ABB56319	Pyrodicticum sp. es
38	41	46.6	955	18	AAW31363	Non-endogenous hum
39	40	45.5	108	22	AAW52456	Cell membrane prot
40	40	45.5	182	23	ABB97829	Mycobacterium tube
41	40	45.5	201	22	ABG27232	Human secretory po
42	40	45.5	201	22	AAW39656	Novel human diago
43	40	45.5	201	22	AAW95306	Human polypeptide
44	40	45.5	215	22	AAW41442	Human protein sequ
45	40	45.5	220	13	AAW21500	Human polypeptide

ALIGNMENTS

RESULT 1	ABG13124	standard; Protein; 368 AA.
ID	ABG13124	
XX	ABG13124:	
AC	18-FEB-2002	(first entry)
DT	XX	
DE	Novel human diagnostic protein #13115.	
XX	XX	
KW	Human: chromosome mapping; gene mapping; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder.	
XX	XX	
OS	Homo sapiens.	
XX	XX	
PN	WO200175067-A2.	
XX	XX	
PD	11-OCT-2001.	
XX	XX	
PF	30-MAR-2001; 2001WO-US08631.	
XX	XX	
PR	31-MAR-2000; 2000US-0540217.	
PR	23-AUG-2000; 2000US-0649167.	
XX	XX	
PA	(HYSE-) HYSEQ INC.	
PI	Drmanac RT, Liu C, Tang YF;	
XX	XX	
DR	WPI: 2001-639362/73.	
DR	N-PSDB; AAS7311.	
XX	XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in	
PT	diagnostics, forensics, gene mapping, identification of mutations	

PT		responsible for genetic disorders or other traits and to assess biodiversity -
PS		
XX		
XX		
PS		Claim 20; SEQ ID No 43483; 103pp; English.
CC		The invention relates to isolated polynucleotide (I) and
CC		polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC		polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC		and gene mapping, and in recombinant production of (II). The
CC		polynucleotides are also used in diagnostics as expressed sequence tags
CC		for identifying expressed genes. (I) is useful in gene therapy techniques
CC		to restore normal activity of (II) or to treat disease states involving
CC		(II). (II) is useful for generating antibodies against it, detecting or
CC		quantitating a polypeptide in tissue, as molecular weight markers and as
CC		a food supplement. (II) and its binding partners are useful in medical
CC		imaging of sites expressing (II). (I) and (II) are useful for treating
CC		disorders involving aberrant protein expression or biological activity.
CC		The polypeptide and polynucleotide sequences have applications in
CC		diagnostics, forensics, gene mapping, identification of mutations
CC		responsible for genetic disorders or other traits to assess biodiversity
CC		and to produce other types of data and products dependent on DNA and
CC		amino acid sequences. ABC00010-ABC03077 represent novel human
CC		diagnostic amino acid sequences of the invention.
CC		Note: The sequence data for this patent did not appear in the printed
CC		specification, but was obtained in electronic format directly from Wipo
CC		at ftp.wipo.int/pub/published_pcl_sequences.
SO		
Sequence	368 AA;	
Query Match	54.5%; Score 48; DB 22; Length 368;	
Best Local Similarity	76.9%; Pred. No. 6.9;	
Matches	10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
Oy	6 VLAIVGVAVIPAH 18 :	
Dd	145 VLALGLGAVAPAH 157	
RESULT 2		
AAM10017		
ID	AAM10017 standard; Protein: 242 AA.	
XX		
AC	AAM10017;	
XX		
DT	16-SEP-1997 (first entry)	
XX		
DE	G-protein coupled receptor.	
XX		
KW	receptor; human; brain; recombinant production; ligand; screening; antibody; gene therapy.	
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Protein	10..233
FT	/note= "claim 1"	
XX		
FN	JF09121865-A.	
XX		
PB	13-MAY-1997.	
XX		
PF	30-OCT-1995; 95JP-0303301.	
XX		
PR	30-OCT-1995; 95JP-0303301.	
XX		
PA	(TAKE) TAKEDA CHEM IND LTD.	
XX		
DR	WPI; 1997-314225/29.	
XX		
DR	N-PSDB; AAT77788.	
XX		
PT	DNA encoding G protein-coupled receptor protein - used to screen for	
PT	ligands and in treatment of disease	
XX		

PS Claim 1; Fig 4; 34pp: Japanese.

XX This sequence is a G-protein-coupled receptor protein encoded by cDNA
CC (AA777788) synthesised from human brain-derived polyA+ RNA. The cDNA,
CC vectors and host cells are used for the recombinant production of the
CC protein receptor. The receptor and cDNA are also used to determine a
CC ligand against the receptor, for generation of antibodies and for gene
CC therapy, etc..

XX Sequence 242 AA:

SO Query Match 52.3%; Score 46; DB 18; Length 242;
Best Local Similarity 66.7%; Pred. No. 9.4;

Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0

OY 2 KVSRYLALVLCGVIP 16
| | | | | | | | | |
Db 132 KASRYTTLVLGFVLP 146

RESULT 3

ID AAR72985 standard; Protein; 333 AA.
XX AAR72985;
XX AAR72985;
XX 04-NOV-1995 (first entry)
XX Epsilon opioid receptor.
DE Epsilon opioid receptor.
XX human; epsilon opioid receptor; analgesic; drug screening;
KW drug design.
XX Homo sapiens.
OS
PN W09512670-A.
PD 11-MAY-1995.
XX
PF 07-NOV-1994; 94MO-CA00613.
PR 05-NOV-1993; 93US-0148215.
XX
PA (ALCO-) ALCOHOLISM & DRUG ADDICTION RES FOUND.
PI O'Down BF;
XX
DR MPI: 1995-185775/24.
DR N-PDB: AAQ83682.
XX
PT New polynucleotide encoding an epsilon opioid receptor polypeptide -
PS useful for design of drugs, e.g. analgesics
XX
PS Claim 3; Page 91-93; 114pp: English.
XX

The sequence represents a human epsilon opioid receptor, encoded by
a DNA sequence from clone-11. The protein may be produced in
recombinant form in host cells for use in design and screening of
analgesic drugs for enhanced specificity of function.

XX Sequence 333 AA:
SO Query Match 52.3%; Score 46; DB 16; Length 333;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0.

OY 2 KVSRYLALVLCGVIP 16
| | | | | | | | | |
Db 210 KASRYTTLVLGFVLP 224

RESULT 4
KAY90613

ID AAY90613 standard; Protein: 333 AA.
XX
XX AAY90613;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human G protein-coupled receptor GPR8.
XX
XX G protein-coupled receptor; GPCR; constitutively active;
KW intracellular loop 3; transmembrane domain 6; drug screening;
KW agonist; antagonist.
XX
OS Homo sapiens.
XX
PN MO200022129-A1.
XX
PD 20-APR-2000.
XX
PF 12-OCT-1999; 99WO-US33938.
XX
PR 13-OCT-1998; 98US-0170496.
XX
PA (AREN-) ARENA PHARM INC.
XX
PI Behan DP, Chalmers DT, Liaw CW;
XX
DR WPI: 2000-329165/28.
XX
DR N-PSDB; AAA30590.
XX
PT Non-endogenous constitutively activated human G protein-coupled
PT receptors, useful for identifying agonists for use as pharmaceutical
PT agents
XX
PS Example 1; Page 113-114; 341pp: English.
XX
XX The invention relates to constitutively active, non-endogenous versions
CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC and AAA30775-A30779). The mutant proteins of the invention contain a
CC mutation in a portion of the protein comprising intracellular loop 3
CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC is substituted for an endogenous residue in IC3 at a position 16 amino
CC acids N-terminal of an endogenous proline in TM6 to form a sequence
CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC or Ala, and is preferably Lys. When the endogenous residue at this
CC position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC The 15 amino acid stretch between the substituted amino acid and the Pro
CC may be endogenous, non-endogenous, or a mixture of endogenous and
CC non-endogenous residues. The constitutively active GPCRs are useful for
CC identifying antagonists, agonists and partial agonists for use as
CC pharmaceutical agents. The mutant proteins are also useful in research
CC settings for elucidating the roles of the receptors in normal and
CC diseased conditions. Antagonists for a particular GPCR are useful for
CC treating diseases and disorders associated with that receptor. Because
CC the novel mutant GPCRs are constitutively active, they can be used
CC directly for screening of compounds without the need for endogenous
CC ligands. The present sequence represents a human wild-type GPCR referred
CC to in an exemplification of the invention.
XX
SQ Sequence 333 AA:
Query Match 52.3%; Score 46; DB 21; Length 333;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 KVSRYLALVLAIVIP 16
I I I I I I I I I I
Db 210 KASRYTVTLVGLFVLP 224

RESULT 5
AAY90647
ID AAY90647 standard; Protein: 333 AA.

XX
AC AAY90647;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human mutant G protein-coupled receptor GPR8 (T259K).
XX
XX G protein-coupled receptor; GPCR; constitutively active;
KW intracellular loop 3; transmembrane domain 6; drug screening;
KW agonist; antagonist; mutant; mutein.
XX
OS Homo sapiens.
XX
PN MO200022129-A1.
XX
PD 20-APR-2000.
XX
PF 12-OCT-1999; 99WO-US33938.
XX
PR 13-OCT-1998; 98US-0170496.
XX
PA (AREN-) ARENA PHARM INC.
XX
PI Behan DP, Chalmers DT, Liaw CW;
XX
DR WPI: 2000-329165/28.
XX
DR N-PSDB; AAA30713.
XX
PT Non-endogenous constitutively activated human G protein-coupled
PT receptors, useful for identifying agonists for use as pharmaceutical
PT agents
XX
PS Example 2; Page 221-222; 341pp: English.
XX
XX The invention relates to constitutively active, non-endogenous versions
CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
CC and AAA30775-A30779). The mutant proteins of the invention contain a
CC mutation in a portion of the protein comprising intracellular loop 3
CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
CC is substituted for an endogenous residue in IC3 at a position 16 amino
CC acids N-terminal of an endogenous proline in TM6 to form a sequence
CC X-(AA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg
CC or Ala, and is preferably Lys. When the endogenous residue at this
CC position is Lys, this residue is replaced by His, Arg or preferably Ala.
CC The 15 amino acid stretch between the substituted amino acid and the Pro
CC may be endogenous, non-endogenous, or a mixture of endogenous and
CC non-endogenous residues. The constitutively active GPCRs are useful for
CC identifying antagonists, agonists and partial agonists for use as
CC pharmaceutical agents. The mutant proteins are also useful in research
CC settings for elucidating the roles of the receptors in normal and
CC diseased conditions. Antagonists for a particular GPCR are useful for
CC treating diseases and disorders associated with that receptor. Because
CC the novel mutant GPCRs are constitutively active, they can be used
CC directly for screening of compounds without the need for endogenous
CC ligands. Sequences AAY90643- AAY90677 and AAY90683-Y90687 the mutant
CC human GPCRs of the invention.
XX
SQ Sequence 333 AA:
Query Match 52.3%; Score 46; DB 21; Length 333;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 KVSRYLALVLAIVIP 16
I I I I I I I I I I
Db 210 KASRYTVTLVGLFVLP 224

RESULT 6
AAU01295
ID AAU01295 standard; protein: 333 AA.


```

XX AC AAU01295;
XX XX 29-AUG-2001 (first entry)
XX DT
XX DE Human G-protein receptor 8, GPR 8.
XX XX
XX XX G-protein receptor 8; GPR8; phylogenetic profiling; orphan receptor;
XX KM constitutive activity; agonist.
XX OS Homo sapiens.
XX PN WO200127632-A2.
XX PD 19-APR-2001.
XX PF 02-OCT-2000; 2000WO-IB01407.
XX PR 02-OCT-2000; 2000WO-IB01407.
XX PA (CAMP-) CAMBRIDGE DRUG DISCOVERY LTD.
XX PA (WILL/) WILLIAMS K M.
XX PI Wenham D, Packer JC.
XX DR WPI: 2001-282100/29.
XX PT Predicting mutants that alter the activity of receptors using multiple
XX PT sequence alignment and phylogenetic profiling, useful e.g. for altering
XX PT the activities of orphan proteins -
XX PS Example 1; Fig 2; 91pp; English.
XX CC The sequence represents a Human G-protein receptor 8, GPR8. The
XX CC sequence is used in methods of predicting mutations that alter the
XX CC activity of a receptor in a desired manner, comprising utilising
XX CC multiple sequence alignment and phylogenetic profiling to identify the
XX CC relatives of a given receptor that are most likely to provide useful data
XX CC allowing prediction of sites to mutate in the given receptor. The methods
XX CC are applicable to any type of receptor, and are particularly well suited
XX CC for predicting sites to mutate in order to alter the activities of orphan
XX CC receptors for which no agonists are known. In particular, the method is
XX CC used to predict cellular receptor mutations that induce the receptor to
XX CC constitutively activate it's downstream signalling activities.
XX SQ Sequence 333 AA:
XX
XX Query Match 52.3%; Score 46; DB 22; Length 333;
XX Best Local Similarity 66.7%; Pred. No. 13;
XX Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
OY 2 KVSRYLALVGLGVIP 16
DB 210 KASRYTTLVGLGVLP 224

```

```

FT Misc-difference 124
XX /note= "Wild-type Asp substituted by Ala"
XX FT
XX PN WO200127632-A2.
XX PD 19-APR-2001.
XX PF 02-OCT-2000; 2000WO-IB01407.
XX PR 02-OCT-2000; 2000WO-IB01407.
XX PA (CAMP-) CAMBRIDGE DRUG DISCOVERY LTD.
XX PA (WILL/) WILLIAMS K M.
XX PI Wenham D, Packer JC.
XX DR WPI: 2001-282100/29.
XX PT Predicting mutants that alter the activity of receptors using multiple
XX PT sequence alignment and phylogenetic profiling, useful e.g. for altering
XX PT the activities of orphan proteins -
XX PS Claim 53; Page - : 91pp; English.
XX CC The sequence represents a Human G-protein receptor 8, GPR8, mutant
XX CC D124A, predicted by the method of the invention to be a constitutively
XX CC active mutant. The sequence is used in methods of predicting mutations
XX CC that alter the activity of a receptor in a desired manner, comprising
XX CC utilising multiple sequence alignment and phylogenetic profiling to
XX CC identify the relatives of a given receptor that are most likely to
XX CC provide useful data allowing prediction of sites to mutate in the given
XX CC receptor. The methods are applicable to any type of receptor, and are
XX CC particularly well suited for predicting sites to mutate in order to alter
XX CC the activities of orphan receptors for which no agonists are known. In
XX CC particular, the method is used to predict cellular receptor mutations
XX CC that induce the receptor to constitutively activate it's downstream
XX CC signalling activities.
XX CC Note: The present sequence is not shown in the specification but is
XX CC derived from the GPR8 sequence shown in AAU01295.
XX SQ Sequence 333 AA:
XX
XX Query Match 52.3%; Score 46; DB 22; Length 333;
XX Best Local Similarity 66.7%; Pred. No. 13;
XX Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
OY 2 KVSRYLALVGLGVIP 16
DB 210 KASRYTTLVGLGVLP 224

```

```

RESULT 7
AAU01296
ID AAU01296 standard; protein: 333 AA.
XX
XX AAU01296;
XX
XX 29-AUG-2001 (first entry)
XX
XX Human G-protein receptor 8, GPR 8, mutant D124A.
XX
XX G-protein receptor 8; GPR8; phylogenetic profiling; orphan receptor;
XX KM constitutive activity; agonist; mutant; mutein; D124A.
XX OS Homo sapiens.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX

```

```

RESULT 8
AAU01297
ID AAU01297 standard; protein: 333 AA.
XX
XX AAU01297;
XX
XX 29-AUG-2001 (first entry)
XX
XX Human G-protein receptor 8, GPR 8, mutant N127A.
XX
XX G-protein receptor 8; GPR8; phylogenetic profiling; orphan receptor;
XX KM constitutive activity; agonist; mutant; mutein; N127A.
XX OS Homo sapiens.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 127
XX /note= "Wild-type Asn substituted by Ala"
XX
XX WO200127632-A2.
XX

```

PD 19-APR-2001.
XX
XX 02-OCT-2000; 2000MO-IB01407.
XX
XX 02-OCT-2000; 2000MO-IB01407.
PR
XX (CAMP-) CAMBRIDGE DRUG DISCOVERY LTD.
PA (WILL/) WILLIAMS K M.
PI Wenham D, Packer JC;
XX WPI; 2001-282100/29.
DR
XX
XX Predicting mutants that alter the activity of receptors using multiple
PT sequence alignment and phylogenetic profiling, useful e.g. for altering
PT the activities of orphan proteins -
XX
XX
PS Claim 53; Page - : 91pp; English.
XX
XX The sequence represents a Human G-protein receptor 8, GPR8, mutant
CC N127A, predicted by the method of the invention to be a constitutively
CC active mutant. The sequence is used in methods of predicting mutations
CC that alter the activity of a receptor in a desired manner, comprising
CC utilising multiple sequence alignment and phylogenetic profiling to
CC identify the relatives of a given receptor that are most likely to
CC provide useful data allowing prediction of sites to mutate in the given
CC receptor. The methods are applicable to any type of receptor, and are
CC particularly well suited for predicting sites to mutate in order to alter
CC the activities of orphan receptors for which no agonists are known. In
CC particular, the method is used to predict cellular receptor mutations
CC that induce the receptor to constitutively activate it's downstream
CC signalling activities.
CC Note: The present sequence is not shown in the specification but is
CC derived from the GPR8 sequence shown in AAU01295.
XX
XX Sequence 333 AA;
SQ
Query Match 52.3%; Score 46; DB 22; Length 333;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 KVSRLVLYLGAVIP 16
DB 210 KASRYVTLVIGFVLP 224
RESULT 9
AAU01298
ID AAU01298 standard; protein: 333 AA.
AC AAU01298;
XX
XX 29-AUG-2001 (first entry)
DT
XX Human G-protein receptor 8, GPR 8, mutant T259E.
DE
XX G-protein receptor 8; GPR8; phylogenetic profiling; orphan receptor;
KM constitutive activity; agonist; mutant; mutain; T259E.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX
XX Key Location/Qualifiers
FH Misc-difference 259 /note="Wild-type Thr substituted by Glu"
FT
XX
XX MO200127632-A2.
PN
XX 19-APR-2001.
PD
XX 02-OCT-2000; 2000MO-IB01407.
PF
XX 02-OCT-2000; 2000MO-IB01407.
PR

XX
XX (CAMP-) CAMBRIDGE DRUG DISCOVERY LTD.
PA (WILL/) WILLIAMS K M.
XX
XX Wenham D, Packer JC;
XX WPI; 2001-282100/29.
DR
XX
XX Predicting mutants that alter the activity of receptors using multiple
PT sequence alignment and phylogenetic profiling, useful e.g. for altering
PT the activities of orphan proteins -
XX
XX
PS Claim 53; Page - : 91pp; English.
XX
XX The sequence represents a Human G-protein receptor 8, GPR8, mutant
CC T259E, predicted by the method of the invention to be a constitutively
CC active mutant. The sequence is used in methods of predicting mutations
CC that alter the activity of a receptor in a desired manner, comprising
CC utilising multiple sequence alignment and phylogenetic profiling to
CC identify the relatives of a given receptor that are most likely to
CC provide useful data allowing prediction of sites to mutate in the given
CC receptor. The methods are applicable to any type of receptor, and are
CC particularly well suited for predicting sites to mutate in order to alter
CC the activities of orphan receptors for which no agonists are known. In
CC particular, the method is used to predict cellular receptor mutations
CC that induce the receptor to constitutively activate it's downstream
CC signalling activities.
CC Note: The present sequence is not shown in the specification but is
CC derived from the GPR8 sequence shown in AAU01295.
XX
XX Sequence 333 AA;
SQ
Query Match 52.3%; Score 46; DB 22; Length 333;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 KVSRLVLYLGAVIP 16
DB 210 KASRYVTLVIGFVLP 224
RESULT 10
ABG65918
ID ABG65918 standard; protein: 333 AA.
XX
XX ABG65918;
AC
XX
XX 28-AUG-2002 (first entry)
DT
XX
XX G protein-coupled receptor related peptide #6.
DE
XX Human; rat; receptor: G protein-coupled receptor; anorectic; anabolic;
KM obesity; appetite enhancement; prolactin production; eating disorder;
KM pig; mouse.
XX
XX Homo sapiens.
OS
XX
XX MO200244368-A1.
PN
XX 06-JUN-2002.
PD
XX
XX 29-NOV-2001; 2001WO-JP10418.
PF
XX
XX 30-NOV-2000; 2000JP-0364801.
PR 26-MAR-2001; 2001JP-0087482.
PR 15-MAY-2001; 2001JP-0145434.
PR 06-SEP-2001; 2001JP-0270838.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX
XX Terao Y, Shintani Y, Harada M, Shimomura Y, Mori M;
PI WPI; 2002-471832/50.
XX
XX

XX	New rat and mouse brain-originated G protein-coupled receptor proteins
PT	TC826, useful in diagnosis and developing drugs for prevention or
PT	treatment of obesity or an eating disorder
XX	
PS	Disclosure: Page 235-237; 312pp; Japanese.
XX	
CC	The invention relates to G protein-coupled receptor proteins and their
CC	associated nucleic acids. The sequences are used in diagnosis of diseases
CC	relating to function of the protein and can be used for creating obesity,
CC	enhancing appetite or their salts that can alter binding of the G
CC	protein-coupled receptors. The proteins and encoded DNAs are useful in
CC	diagnosis of and developing drugs for prevention or treatment of obesity
CC	and eating disorders. Sequences ABG65911-ABG65974 represent G
CC	protein-coupled receptor proteins and related peptides.
XX	
SO	Sequence 333 AA:
Query Match	52.3%; Score 46; DB 23; Length 333;
Best Local Similarity	66.7%; Pred. No. 13;
Matches 10; Conservative	1; Mismatches 4; Indels 0; Gaps 0;
OY	2 KVSRLVLYLGVIP 16
	I I I I I I I I I I
Db	210 KASRYVTLVLFGLP 224
RESULT 11	
ABB84683	
ID	ABB84683 standard; Protein: 333 AA.
XX	
AC	ABB84683;
XX	
DT	13-MAY-2002 (first entry)
XX	
DE	Human GPR8-1ligand related protein #1.
XX	
KW	Anorectic; GPR8 ligand; central nervous system; obesity;
RW	appetite-stimulating agent; prolactin; human.
XX	
OS	Homo sapiens.
XX	
PN	WO200198494-A1.
XX	
PD	27-DEC-2001.
XX	
XX	20-JUN-2001; 2001WO-JP05257.
XX	
PF	21-JUN-2000; 2000JP-0191089.
XX	
PR	06-SEP-2000; 2000JP-0275013.
XX	
PR	13-APR-2001; 2001JP-0116000.
XX	
PA	(TAKE) TAKEDA CHEM IND LTD.
XX	
PI	Mori M, Shimomura Y, Harada M, Kurihara M, Kitada C, Asami T;
PI	Matsunoto Y, Adachi Y, Watanabe T, Sugo T, Abe M;
XX	
XX	WPI: 2002-139790/18.
DR	N-PSDB; ABL61580.
XX	
PT	Ligand to GPR8 and encoded gene, useful in developing receptor-binding
PT	assay system, diagnosis and screening candidate compounds for central
PT	nervous system function-regulating drugs to treat e.g. obesity
XX	
PS	Claim 1: Page 167-169; 221pp; Japanese.
XX	
CC	The present invention relates to GPR8 ligands. The ligands as well as
CC	their precursor proteins and DNAs are useful in developing
CC	receptor-binding assay systems, diagnosis and screening candidate
CC	compounds for central nervous system function-regulating drugs as
CC	preventives or remedies for obesity, appetite-stimulating agents and
CC	prolactin production promoters or inhibitors. The present sequence

```

CC      was used to illustrate the invention.
XX
SQ      Sequence      333 AA;
Query Match      52.3%;      Score 46;      DB 23;      Length 333;
Best Local Similarity      66.7%;      Pred. No. 13;
Matches      10;      Conservative      1;      Mismatches      4;      Indels      0;      Gaps      0;
QY      2 KYSRYLALVLGAVIP 16
      1 111 1111 1:1
DB      210 KASRYTTLVLGFLVP 224

RESULT 12
ABB84723
ID      ABB84723 standard; Protein; 333 AA.
XX
AC      ABB84723;
XX
DT      13-MAY-2002 (first entry)
XX
DE      Human GPR8-ligand related protein #3.
XX
KM      Anorectic; GPR8 ligand; central nervous system; obesity;
XX      appetite-stimulating agent; prolactin; human.
XX
OS      Homo sapiens.
XX
PN      WO200198494-A1.
XX
PD      27-DEC-2001.
XX
PF      20-JUN-2001; 2001WO-JP05257.
XX
PR      21-JUN-2000; 2000JP-0191089.
XX      06-SEP-2000; 2000JP-0275013.
XX      13-APR-2001; 2001JP-0116000.
XX
PA      (TAKE ) TAKEDA CHEM IND LTD.
XX
PI      Mori M, Shimomura Y, Harada M, Kurihara M, Kitada C, Asami T;
PI      Matsumoto Y, Adachi Y, Watanabe T, Sugo T, Abe M;
DR      WPI: 2002-139790/18.
DR      N-PSDB; ABL61663.
XX
XX      Ligand to GPR8 and encoded gene, useful in developing receptor-binding
XX      assay system, diagnosis and screening candidate compounds for central
XX      nervous system function-regulating drugs to treat e.g. Obesity
XX
XX      Example 2; Fig 1; 221pp; Japanese.
XX
XX      The present invention relates to GPR8 ligands. The ligands as well as
XX      their precursor proteins and DNAs are useful in developing
XX      receptor-binding assay systems, diagnosis and screening candidate
XX      compounds for central nervous system function-regulating drugs as
XX      preventives or remedies for obesity, appetite-stimulating agents and
XX      prolactin production promoters or inhibitors. The present sequence
XX      was used to illustrate the invention.
XX
SQ      Sequence      333 AA;
Query Match      52.3%;      Score 46;      DB 23;      Length 333;
Best Local Similarity      66.7%;      Pred. No. 13;
Matches      10;      Conservative      1;      Mismatches      4;      Indels      0;      Gaps      0;
QY      2 KYSRYLALVLGAVIP 16
      1 111 1111 1:1
DB      210 KASRYTTLVLGFLVP 224

RESULT 13
ABG28734

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: April 26, 2003, 13:21:37 ; Search time 0.590203 Seconds
(without alignments)
947.192 Million cell updates/sec

Title: US-10-026-994-3

Perfect score: 88

Sequence: 1 MKVSRVLALVGVIPAH 19

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	52.3	333	US-08-148-215A-4	Sequence 4, Appl
2	43.5	48.4	776	US-09-346-237-4	Sequence 4, Appl
3	43.5	48.4	776	US-09-346-237-7	Sequence 7, Appl
4	43	48.9	233	US-09-040-483-4	Sequence 4, Appl
5	42	47.7	365	US-08-370-542-2	Sequence 2, Appl
6	42	47.7	365	US-08-542-358-2	Sequence 2, Appl
7	42	47.7	365	US-08-157-185-16	Sequence 16, Appl
8	42	47.7	365	US-08-281-526B-16	Sequence 16, Appl
9	42	47.7	365	US-09-018-351-2	Sequence 2, Appl
10	42	47.7	365	US-09-450-790A-16	Sequence 16, Appl
11	42	47.7	365	US-09-332-837-16	Sequence 16, Appl
12	41	46.6	346	US-08-602-359A-34	Sequence 34, Appl
13	40	45.5	222	US-09-040-483-3	Sequence 3, Appl
14	40	45.5	222	US-08-832-488-11	Sequence 11, Appl
15	40	45.5	222	5185431-4	Patent No. 5185431
16	40	45.5	234	US-09-040-483-5	Sequence 5, Appl
17	39.5	44.9	772	US-08-410-784A-5	Sequence 5, Appl
18	39	44.3	21	US-09-227-357-507	Sequence 507, App
19	39	44.3	186	US-09-227-357-507	Sequence 507, App
20	39	44.3	281	US-09-404-258-16	Sequence 16, Appl
21	39	44.3	463	US-08-162-402B-9	Sequence 9, Appl
22	38	43.2	62	US-07-645-029-7	Sequence 7, Appl
23	38	43.2	100	US-08-464-517-10	Sequence 10, Appl
24	38	43.2	100	US-08-246-361A-10	Sequence 10, Appl
25	38	43.2	100	US-08-463-772-10	Sequence 10, Appl
26	38	43.2	100	PCT-US93-05000-10	Sequence 10, Appl
27	38	43.2	106	US-08-464-517-26	Sequence 26, Appl

28	38	43.2	106	US-08-246-361A-26	Sequence 26, Appl
29	38	43.2	106	US-08-463-772-26	Sequence 26, Appl
30	38	43.2	116	PCT-US93-05000-26	Sequence 26, Appl
31	38	43.2	148	US-08-460-694-6	Sequence 6, Appl
32	38	43.2	148	US-08-460-744-6	Sequence 6, Appl
33	38	43.2	148	US-07-667-711B-6	Sequence 6, Appl
34	38	43.2	171	US-08-193-977-5	Sequence 5, Appl
35	38	43.2	281	US-09-404-258-6	Sequence 6, Appl
36	38	43.2	328	US-08-148-215A-2	Sequence 2, Appl
37	38	43.2	433	US-08-522-166-7	Sequence 7, Appl
38	38	43.2	433	US-08-488-382A-7	Sequence 7, Appl
39	38	43.2	433	US-08-480-912-7	Sequence 7, Appl
40	38	43.2	630	US-08-797-366-3	Sequence 3, Appl
41	38	43.2	630	US-08-956-268-3	Sequence 3, Appl
42	37	42.0	207	US-08-381-881-6	Sequence 6, Appl
43	37	42.0	207	US-09-281-221-6	Sequence 6, Appl
44	37	42.0	307	US-08-605-284B-19	Sequence 19, Appl
45	37	42.0	327	US-08-118-270-29	Sequence 29, Appl

ALIGNMENTS

```
RESULT 1
US-08-148-215A-4
; Sequence 4, Application US/08148215A
; Patent No. 5591602
; GENERAL INFORMATION:
; APPLICANT: O Dowd, Brian F.
; TITLE OF INVENTION: Opioid Receptor: Compositions and Methods
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: 321 No. 5591602th Clark Street, Suite 800
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,215A
; FILING DATE: 05-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5591602thrup, Thomas E.
; REGISTRATION NUMBER: 33,268
; REFERENCE/DOCKET NUMBER: OPTA003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-744-0090
; TELEFAX: 312-755-4489
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-148-215A-4

Query Match      52.3%; Score 46; DB 1; Length 333;
Best Local Similarity 66.7%; Pred. No. 3.4;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      2 KVSRLVGVIPAH 16
      1 111 111 111 111
Db      210 KASRVTLVGVFLP 224

RESULT 2
US-09-346-237-4
```

```

Sequence 4, Application US/09346237A
Patent No. 6265197
GENERAL INFORMATION:
APPLICANT: Bisgaard-Frantzen, Henrik
APPLICANT: Svendsen, Allan
TITLE OF INVENTION: Starch Debranching Enzymes
FILE REFERENCE: 5629.200-US
CURRENT APPLICATION NUMBER: US/09/346,237A
CURRENT FILING DATE: 1999-07-01
EARLIER APPLICATION NUMBER: PA 1998 00868
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: 60/094,353
EARLIER FILING DATE: 1998-07-28
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Fastseq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 776
TYPE: PPT
ORGANISM: Pseudomonas amyloclavata
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(776)
OTHER INFORMATION: Isoamylase
US-09-346-237-4

```

Query Match	49.48;	Score 43.5;	DB 4;	Length 776;
Best Local Similarity	38.58;	Pred. No. 22;		
Matches 10;	Conservative 4;	Mismatches 5;	Indels 7;	Gaps 1

```
QY 1 MKVSRVLALVIGAVI-----PAHA 19
    11::11:111
Db 1 MKCPKILALLGCAYLAGVPAMPAHA 26
```

```

US-09-346-237-7
US-09-346-237-7
Sequence 7, Application US/09346237A
Patent No. 6265197
GENERAL INFORMATION:
APPLICANT: Bisgaard-Prantzen, Henrik
APPLICANT: Svendsen, Allan
TITLE OF INVENTION: Starch Debranching Enzymes
FILE REFERENCE: 5629.200-US
CURRENT APPLICATION NUMBER: US/09/346, 237A
CURRENT FILING DATE: 1999-07-01
EARLIER APPLICATION NUMBER: PA 1998 00868
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: 60/094,353
EARLIER FILING DATE: 1998-07-28
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 776
TYPE: PRT
ORGANISM: Pseudomonas species SMP1
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(776)
OTHER INFORMATION: Isocamyase
US-09-346-237-7

```

Query Match	49.48;	Score 43.5;	DB 4;	Length 776;
Best Local Similarity	38.58;	Pred. No. 22;		
Matches 10;	Conservative 4;	Mismatches 5;	Indels 7;	Gaps 1

```
QY      1 MKVSRVLALVLGAVI-----PAHA 19
        ||::||:|||
Db      1 MKCPKILALLGCAVLGVPAMPANA 26
```

RESULT 4
US-09-040-483-4
; Sequence 4, Application US/09040483

1 Patent No. 6143867
2
3 GENERAL INFORMATION:
4
5 APPLICANT: Akribiom, Ingrid E.
6
7 TITLE OF INVENTION: NOVEL HUMAN EOSINOPHIL
8
9 TITLE OF INVENTION: DERIVED BASIC PROTEIN
10
11 NUMBER OF SEQUENCES: 5
12
13 CORRESPONDENCE ADDRESS:
14
15 ADDRESS: Incyte Pharmaceuticals, Inc.
16
17 STREET: 3174 Potter Drive
18
19 CITY: Palo Alto
20
21 STATE: CA
22
23 COUNTRY: US
24
25 ZIP: 94304
26
27 COMPUTER READABLE FORM:
28
29 MEDIUM TYPE: Diskette
30
31 COMPUTER: IBM Compatible
32
33 OPERATING SYSTEM: DOS
34
35 SOFTWARE: FASTSEQ Version 1.5
36
37 CURRENT APPLICATION DATA:
38
39 APPLICATION NUMBER: US/09/040,483
40
41 FILING DATE:
42
43 PRIOR APPLICATION DATA:
44
45 APPLICATION NUMBER: 08/740,036
46
47 FILING DATE:
48

PRIOR APPLICATION DATA: 08/740,036
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0133 US
TELECOMMUNICATION INFORMATION:

TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:

```

: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: IMMEDIATE SOURCE:
: LIBRARY: GenBank
: CLONE: 220291
:
: JS-09-040-483-4

```

Query Match	48.98;	Score 43;	DB 4;	Length 233;
Best Local Similarity	55.68;	Pred. No. 7.3;		
Matches 10; Conservative	3;	Mismatches 5;	Indels 0;	Gaps 0;

```
QY      1 MKVSRVIALVLGAVIPAH 18
        ||: :||: ||| |
Db      1 MKLLLLALLLGAVSTRH 18
```

RESULT 5
 US-08-370-542-2
 Sequence 2, Application US/08370542
 Patent No. 5476782
 GENERAL INFORMATION:
 APPLICANT: Weinstank, Richard L.
 APPLICANT: Branchek, Theresa
 APPLICANT: Hartly, Paul R.
 TITLE OF INVENTION: DNA ENCODING A HUMAN 5-TITLE RECEPTOR AND
 TITLE OF INVENTION: USES THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham
 STREET: 30 Rockefeller Plaza
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10112
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,542
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/194,113
FILING DATE:
APPLICATION NUMBER: US/07/803,626
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1795/39317
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-370-542-2

Query Match
Best Local Similarity 47.7%; Score 42; DB 1; Length 365;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 KVSRLVLYGAVI 15
DB 288 KAARILGLIGAFI 301

RESULT 6
US-08-542-358-2
Sequence 2, Application US/08542358
Patent No. 5786155
GENERAL INFORMATION:
APPLICANT: Weinshtank, Richard L.
APPLICANT: Branchek, Theresa
APPLICANT: Hartig, Paul R.
TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1E RECEPTOR AND USES THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/542,358
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1795/39317-22/JPW/MAT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-542-358-2

Query Match
Best Local Similarity 47.7%; Score 42; DB 1; Length 365;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 KVSRLVLYGAVI 15
DB 288 KAARILGLIGAFI 301

RESULT 7
US-08-157-185-16
Sequence 16, Application US/08157185
Patent No. 5985585
GENERAL INFORMATION:
APPLICANT: Bard A. Jonathan
APPLICANT: Branchek A. Theresa
APPLICANT: Weinshtank L. Richard
TITLE OF INVENTION: DNA ENCODING A HUMAN SEROTONIN
RECEPTOR (5-HT4B) AND USES THEREOF
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/157,185
FILING DATE: 15-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White P., John
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41908-A-PCT-US/JPW/MAT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-157-185-16

Query Match
Best Local Similarity 47.7%; Score 42; DB 2; Length 365;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 KVSRLVLYGAVI 15
DB 288 KAARILGLIGAFI 301

RESULT 8
US-08-281-526B-16
Sequence 16, Application US/08281526B
Patent No. 6083749
GENERAL INFORMATION:
APPLICANT: Bard A. Jonathan
APPLICANT: Branchek A. Theresa
APPLICANT: Weinshtank L. Richard

TITLE OF INVENTION: DNA ENCODING A HUMAN SEROTONIN
RECEPTOR (5-HT4B) AND USES THEREOF
TITLE OF INVENTION: RECEPTOR (5-HT4B) AND USES THEREOF
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOPER & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/281,526B
FILING DATE: 27-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White P., John
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41908-1/JPM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-281-526B-16

Query Match 47.7%; Score 42; DB 3; Length 365;
Best Local Similarity 57.1%; Pred. No. 18;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 KVSRLVLYGAVI 15
DB 288 KAARILGLIGAFI 301

RESULT 9
US-09-018-351-2
Sequence 2, Application US/09018351
Patent No. 6096507
GENERAL INFORMATION:
APPLICANT: Weinschank, Richard L.
APPLICANT: Branchek, Theresa
TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1E RECEPTOR
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,351
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/542,358

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1795/39317-22/JPM/MAT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-018-351-2

Query Match 47.7%; Score 42; DB 3; Length 365;
Best Local Similarity 57.1%; Pred. No. 18;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 KVSRLVLYGAVI 15
DB 288 KAARILGLIGAFI 301

RESULT 10
US-09-450-790A-16
Sequence 16, Application US/09450790A
Patent No. 6376243
GENERAL INFORMATION:
APPLICANT: Bard, Jonathan A
APPLICANT: Branchek, Theresa
TITLE OF INVENTION: DNA ENCODING A HUMAN SEROTONIN RECEPTOR (5HT4B) AND USES THERE
FILE REFERENCE: 1795/419081A
CURRENT APPLICATION NUMBER: US/09/450,790A
CURRENT FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: 08/281,526
PRIOR FILING DATE: 1994-07-27
PRIOR APPLICATION NUMBER: 07/971,690
PRIOR FILING DATE: 1992-11-03
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.1
SEQ ID NO 16
LENGTH: 365
TYPE: PRT
ORGANISM: Homo Sapiens
US-09-450-790A-16

Query Match 47.7%; Score 42; DB 4; Length 365;
Best Local Similarity 57.1%; Pred. No. 18;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 KVSRLVLYGAVI 15
DB 288 KAARILGLIGAFI 301

RESULT 11
US-09-332-837-16
Sequence 16, Application US/09332837
Patent No. 6432655
GENERAL INFORMATION:
APPLICANT: Bard, Jonathan
APPLICANT: Branchek, Theresa
APPLICANT: Weinschank, Richard
TITLE OF INVENTION: Methods of Obtaining Pharmaceutical Compositions
FILE REFERENCE: 41908-AA-PCT-US
CURRENT APPLICATION NUMBER: US/09/332,837
CURRENT FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.0
SEQ ID NO 16

```

; LENGTH: 365
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-332-837-16

```

Query Match	47.7%	Score 42;	DB 4;	Length 365;
Best Local Similarity	57.1%;	Pred. No. 18;		
Matches	8;	Conservative	3;	Mismatches 3;
			Indels	0;
			Gaps	0

```
Qy      2 KYSRVLALVGAVI 15      | : : | : | | |
Db      288 KAARILGLITGAFI 301
```

RESULT 12
US-08-602-359A-34
: Sequence 34, Application US/08602359A
: Patent No. 5942430
GENERAL INFORMATION:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

ADDRESSEE: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA

```
;
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: 3.5 INCH DISKETTE
;
;      COMPUTER: IBM PS/2
;      OPERATING SYSTEM: MS-DOS
;      SOFTWARE: WORD PERFECT 6.0
;
```

;; APPLICATION NUMBER: US/08/602,359A
;; FILING DATE: February 16, 1996

```

;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:

```

;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:

```

;      REGISTRATION NUMBER: 38,347
;      REFERENCE/DOCKET NUMBER: 09010/010001
;      TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099

SEQUENCE CHARACTERISTICS:
LENGTH: 346 AMINO ACIDS
TYPE: AMINO ACID

```

; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-602-359A-34

```

Query Match	46.68;	Score 41;	DB 2;	Length 346;
Best Local Similarity	61.58;	Pred. No. 24;		
Matches	8; Conservative	3; Mismatches	2; Indels	0; Gaps

```
QY      6 VLALVGLGAVIPAH 18
          ||||: |:|:
Db      67 VLALVSVVVPAY 79
```

RESULT 13

US-09-040-483-3
; Sequence 3, Application US/09040483
; Patent No. 6143867

APPLICANT: Akerblom, Ingrid E.
TITLE OF INVENTION: NOVEL HUMAN EOSINOPHIL
DERIVED BASIC PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA

```

?      ZIP: 94304
?
?      COMPUTER READABLE FORM:
?
?      MEDIUM TYPE: Diskette
?
?      COMPUTER: IBM Compatible
?
?      OPERATING SYSTEM: DOS
?
?      SOFTWARE: FastSeq Version 1.5
?
?      CURRENT APPLICATION DATA:
?
?      APPLICATION NUMBER: TS-008/060 ARB

```

```

; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/740,036
; FILING DATE:

```

FILED DATE: _____
ATTORNEY/AGENT INFORMATION: _____

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0133 US
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166

```

; ELEFA: 413-043-4100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 222 amino acids
TYPE: amino acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:

```

```

; LIBRARY: GenBank
; CLONE: 34476

```

US-09-040-483-3

Query Match	45.5%	Score 40;	DB 4;	Length 222;
Best Local Similarity	50.0%;	Pred. NO. 22;		
Matches 9; Conservative		3; Mismatches	6; Indels	0; Gaps 0

```
QY 1 MKVSRVLALVLGAVIPAH 18
    ||: :||: ||| |
Db 1 MKLPLLLALLEGAVSALH 18
```

RESULT 14
US-08-832-488-11
; Sequence 11, Application US/08832488
; Patent No. 6448044
; GENERAL INFORMATION.

APPLICANT: NI, JIAN
APPLICANT: LI, HAOBONG
APPLICANT: YU, GUO-LING
APPLICANT: GENG², REIMER, I.
TITLE OF INVENTION: HUMAN NATURAL KILLER CELL ACTIVATION
TITLE OF INVENTION: FACTOR II
NUMBER OF SEQUENCES: 11
CORRESPONDENCE: 11

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: HUMAN GENOME SCIENCES, INC
3
4 STREET: 9410 KEY WEST AVENUE
5
6 CITY: ROCKVILLE
7
8 STATE: MD
9
10 COUNTRY: US
11
12 ZIP: 20850

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:25:17 ; Search time 0.749104 Seconds
(without alignments)
2032.383 Million cell updates/sec

Title: US-10-026-994-3

Perfect score: 88

Sequence: 1 MKVSRVLAIVLGVAVIPAH 19

Scoring table: BIOSUM62

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43.5	49.4	776	10 US-09-833-435A-4	Sequence 4, Appli
2	43.5	49.4	776	10 US-09-833-435A-7	Sequence 7, Appli
3	42	47.7	114	10 US-09-864-761-41753	Sequence 41753, A
4	41	46.6	346	9 US-10-027-805-34	Sequence 34, Appl
5	41	46.6	346	9 US-10-027-804-34	Sequence 34, Appl
6	41	46.6	346	10 US-09-903-410-34	Sequence 34, Appl
7	40	45.5	314	9 US-10-101-464A-728	Sequence 728, App
8	40	45.5	774	10 US-09-815-242-12046	Sequence 12046, A
9	39	44.3	21	9 US-09-983-802-502	Sequence 502, App
10	39	44.3	186	9 US-09-983-802-507	Sequence 507, App
11	39	44.3	281	9 US-09-853-701-16	Sequence 16, Appl
12	39	44.3	326	10 US-09-925-300-1605	Sequence 1605, Ap
13	39	44.3	861	10 US-09-815-242-14088	Sequence 14088, A
14	38	43.2	130	9 US-09-925-299-1128	Sequence 1128, Ap
15	38	43.2	130	10 US-09-925-299-1128	Sequence 1128, Ap
16	38	43.2	218	9 US-10-142-201B-8	Sequence 8, Appli
17	38	43.2	268	10 US-09-875-456A-14	Sequence 14, Appl
18	38	43.2	281	9 US-09-853-701-6	Sequence 6, Appli
19	38	43.2	379	12 US-10-007-693-105	Sequence 105, App

20	38	43.2	461	12 US-10-007-693-72	Sequence 72, Appl
21	37	42.0	76	10 US-09-864-761-42332	Sequence 42332, A
22	37	42.0	109	10 US-09-867-550-578	Sequence 578, App
23	37	42.0	273	9 US-10-063-547-168	Sequence 168, App
24	37	42.0	273	9 US-10-036-041-18	Sequence 18, Appl
25	37	42.0	273	9 US-10-028-072-540	Sequence 540, App
26	37	42.0	273	9 US-10-035-855-18	Sequence 18, Appl
27	37	42.0	273	9 US-10-063-616-168	Sequence 168, App
28	37	42.0	273	9 US-10-063-502-168	Sequence 168, App
29	37	42.0	273	9 US-10-121-049-540	Sequence 540, App
30	37	42.0	273	9 US-10-123-904-540	Sequence 540, App
31	37	42.0	273	9 US-10-140-470-540	Sequence 540, App
32	37	42.0	273	9 US-09-931-836-18	Sequence 18, Appl
33	37	42.0	273	9 US-10-175-746-540	Sequence 540, App
34	37	42.0	273	9 US-10-176-918-540	Sequence 540, App
35	37	42.0	273	9 US-10-176-921-540	Sequence 540, App
36	37	42.0	273	9 US-10-036-214-18	Sequence 18, Appl
37	37	42.0	273	9 US-10-137-865-540	Sequence 540, App
38	37	42.0	273	9 US-10-140-474-540	Sequence 540, App
39	37	42.0	273	9 US-10-035-719-18	Sequence 18, Appl
40	37	42.0	273	9 US-10-142-431-540	Sequence 540, App
41	37	42.0	273	9 US-10-143-114-540	Sequence 540, App
42	37	42.0	273	9 US-10-140-002-540	Sequence 540, App
43	37	42.0	273	9 US-10-036-160-18	Sequence 18, Appl
44	37	42.0	273	9 US-10-142-419-540	Sequence 540, App
45	37	42.0	273	9 US-10-035-958-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1

US-09-833-435A-4

Sequence 4, Application US/09833435A

Patent No. US20020081670A1

GENERAL INFORMATION:

APPLICANT: Bisgard-Frantzen, Henrik

APPLICANT: Svendsen, Allan

TITLE OF INVENTION: Starch Debranching Enzymes

FILE REFERENCE: 5629.200-US

CURRENT APPLICATION NUMBER: US/09/833,435A

CURRENT FILING DATE: 2001-04-12

PRIOR APPLICATION NUMBER: 09/346,237

PRIOR FILING DATE: 1999-07-01

PRIOR APPLICATION NUMBER: 60/094,353

NUMBER OF SEQ ID NOS: 14

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4

LENGTH: 776

TYPE: PRT

ORGANISM: Pseudomonas amyloclavata

FEATURE:

NAME/KEY: PEPTIDE

LOCATION: (1)..(776)

OTHER INFORMATION: Isoamylase

US-09-833-435A-4

Query Match 49.4%; Score 43.5; DB 10; Length 776;

Best Local Similarity 38.5%; Pred. No. 52;

Matches 10; Conservative 4; Mismatches 5; Indels 7; Gaps 1;

QY 1 MKVSRVLAIVLGVAVIPAH 19

DB 1 MKCPKTLAALGVAVIPAMP 26

RESULT 2

US-09-833-435A-7

Sequence 7, Application US/09833435A

Patent No. US20020081670A1

GENERAL INFORMATION:

APPLICANT: Bisgard-Frantzen, Henrik

APPLICANT: Svendsen, Allan
TITLE OF INVENTION: Starch Debranching Enzymes
FILE REFERENCE: 5629,200-US
CURRENT APPLICATION NUMBER: US/09/833,435A
CURRENT FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 09/346,237
PRIOR FILING DATE: 1999-07-01
PRIOR APPLICATION NUMBER: 60/094,353
PRIOR FILING DATE: 1998-07-28
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 7
LENGTH: 776
TYPE: PRT
ORGANISM: Pseudomonas species SMP1
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(776)
OTHER INFORMATION: Isoamylase
US-09-833-435A-7

Query Match 49.4%; Score 43.5; DB 10; Length 776;
Best Local Similarity 38.5%; Pred. No. 52;
Matches 10; Conservative 4; Mismatches 5; Indels 7; Gaps 1;

OY 1 MKVSRVLAIVLGAVI-----PAHA 19
DB 1 MKCPKILALGCAVLGAVPAMPAHA 26

RESULT 3
US-09-864-761-41753
Sequence 41753, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aemulca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO: 41753
LENGTH: 114
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO A138827.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
OTHER INFORMATION: EST_HUMAN HIT: BF196846.1, EVALUATE 1.00e-21
OTHER INFORMATION: SWISSPROT HIT: P28566, EVALUATE 4.00e-62
US-09-864-761-41753

Query Match 47.7%; Score 42; DB 10; Length 114;
Best Local Similarity 57.1%; Pred. No. 10;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 KVSRLVLAIVLGAVI 15
DB 37 KARILIGLIGAVI 50

RESULT 4
US-10-027-805-34
Sequence 34, Application US/10027805
Patent No. US20020164725A1

GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
MURPHY, Dennis
REID, John
MAFFIA, Anthony
LINK, Steven
SMANSON, Ronald V.
WARREN, Patrick V.
KOSMOTKA, Anna

TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/027,805
FILING DATE: 21-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/602,359
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347

```

: REFERENCE/DOCKET NUMBER: 09010/010001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-678-5070
: TELEFAX: 619-678-5099
: INFORMATION FOR SEQ ID NO: 34:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 346 AMINO ACIDS
: TYPE: AMINO ACID
: TOPOLOGY: LINEAR
: MOLECULE TYPE: PROTEIN
: SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-10-027-805-34

Query Match      46.6%; Score 41; DB 9; Length 346;
Best Local Similarity 61.5%; Pred. No. 53;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 VLALVGVAVIPAH 18
DB 67 VLALVSVVVPAY 79

RESULT 5
US-10-027-804-34
: Sequence 34, Application US/10027804
: Publication No. US20030054530A1
: GENERAL INFORMATION:
: APPLICANT: ROBERTSON, Daniel E.
: MURPHY, Dennis
: REID, John
: MAFFIA, Anthony
: LINK, Steven
: SWANSON, Ronald V.
: WARREN, Patrick V.
: KOSMOTKA, Anna
: TITLE OF INVENTION: ESTERASES
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FISH & RICHARDSON P.C.
: STREET: 4225 EXECUTIVE SQUARE, STE 1400
: CITY: LA JOLLA
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH DISKETTE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WORD PERFECT 6.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/10/027,804
: FILING DATE: 21-Dec-2001
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/602,359
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: HAILE, LISA A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 09010/010001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-678-5070
: TELEFAX: 619-678-5099
: INFORMATION FOR SEQ ID NO: 34:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 346 AMINO ACIDS
: TYPE: AMINO ACID
: TOPOLOGY: LINEAR
: MOLECULE TYPE: PROTEIN
: SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-10-027-804-34

Query Match      46.6%; Score 41; DB 9; Length 346;
```

```

: Best Local Similarity 61.5%; Pred. No. 53;
: Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 VLALVGVAVIPAH 18
DB 67 VLALVSVVVPAY 79

RESULT 6
US-09-903-410-34
: Sequence 34, Application US/09903410
: Patent No. US20020146799A1
: GENERAL INFORMATION:
: APPLICANT: DIVERSA CORPORATION
: APPLICANT: ROBERTSON, Dan
: APPLICANT: MURPHY, Dennis
: APPLICANT: REID, John
: APPLICANT: MAFFIA, Anthony
: APPLICANT: LINK, Steven
: APPLICANT: SWANSON, Ronald
: APPLICANT: WARREN, Patrick
: APPLICANT: KOSMOTKA, Anna
: TITLE OF INVENTION: ENZYMES HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF
: FILE REFERENCE: DIVER1180-2
: CURRENT APPLICATION NUMBER: US/09/903,410
: CURRENT FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: US 09/382,242
: PRIOR FILING DATE: 1999-08-24
: PRIOR APPLICATION NUMBER: US 08/602,359
: PRIOR FILING DATE: 1996-02-16
: NUMBER OF SEQ ID NOS: 42
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 34
: LENGTH: 346
: TYPE: PRT
: ORGANISM: Pyrodicticum
US-09-903-410-34

Query Match      46.6%; Score 41; DB 10; Length 346;
Best Local Similarity 61.5%; Pred. No. 53;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 VLALVGVAVIPAH 18
DB 67 VLALVSVVVPAY 79

RESULT 7
US-10-101-464A-728
: Sequence 728, Application US/10101464A
: Publication No. US20030046728A1
: GENERAL INFORMATION:
: APPLICANT: Strabala, Timothy
: APPLICANT: Nieuwenhuizen, Nicolaas
: APPLICANT: Higgins, Colleen M.
: TITLE OF INVENTION: Compositions Isolated from Plant Cells
: FILE REFERENCE: and their use in the Modification of Plant Cell Signaling
: CURRENT APPLICATION NUMBER: US/10/101,464A
: CURRENT FILING DATE: 2002-03-18
: PRIOR APPLICATION NUMBER: 09/704,302
: PRIOR FILING DATE: 2000-11-01
: PRIOR APPLICATION NUMBER: 09/228,986
: PRIOR FILING DATE: 1999-01-12
: PRIOR APPLICATION NUMBER: 60/162,866
: PRIOR FILING DATE: 1999-11-01
: PRIOR APPLICATION NUMBER: PCT/US00/00724
: PRIOR FILING DATE: 2000-01-11
: NUMBER OF SEQ ID NOS: 989
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 728
: LENGTH: 314
: TYPE: PRT
```



```

: ORGANISM: Pinus radiata
US-10-101-464A-726

Query Match          45.5%; Score 40; DB 9; Length 314;
Best Local Similarity 61.5%; Pred. No. 69;
Matches      8; Conservative           3; Mismatches           2; Indels           0; Gaps           0;

OY       2 KVSRLVALVIGAV 14
         | |:| :||:||||
Db        271 KKSKTLATITLGA 283

RESULT 8
US-09-815-242-12046
: Sequence 12046, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 12046
: LENGTH: 774
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12046

Query Match          45.5%; Score 40; DB 10; Length 774;
Best Local Similarity 38.9%; Pred. No. 1.9e+02;
Matches      7; Conservative           6; Mismatches           5; Indels           0; Gaps           0;

OY       1 MKSVRLALVIGAVPAH 18
         |:|:|: ||::| |||
Db        149 LAVARLEPLIGLYPAH 166

RESULT 9
US-09-983-802-502
: Sequence 502, Application US/09983802
: Publication No. US20030022185A1
: GENERAL INFORMATION:
: APPLICANT: Fischer et al.
: TITLE OF INVENTION: 123 Human Secreted Proteins
: FILE REFERENCE: PZ010P1
: CURRENT APPLICATION NUMBER: US/09/983,802
: PRIOR FILING DATE: 2001-10-25
: PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357
: PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08

```

[illegible]

Query Match 44.3%; Score 39; DB 9; Length 21;
Best Local Similarity 43.8%; Pred. No. 4.8;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Oy 3 VSRVLAIVGAVIPAH 18
Db 6 ISRAVIVFGMLYPAY 21

RESULT 10

US-09-983-802-507
; Sequence 507, Application US/09983802
; Publication No. US20030022185A1
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT FILING DATE: 2001-10-25
; APPLICATION NUMBER: US/09/983,802
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360

; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 507
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (181)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-983-802-507

Query Match 44.3%; Score 39; DB 9; Length 186;
Best Local Similarity 43.8%; Pred. No. 56;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Oy 3 VSRVLAIVGAVIPAH 18
Db 6 ISRAVIVFGMLYPAY 21

RESULT 11

US-09-853-701-16
; Sequence 16, Application US/09853701
; Publication No. US20020182686A1
; GENERAL INFORMATION:
; APPLICANT: YANG, Shuei
; TITLE OF INVENTION: METHOD AND COMPOSITIONS FOR IMPROVED POLYNUCLEOTIDE
; FILE REFERENCE: 089962/0102
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/404,258
; PRIOR FILING DATE: 1999-09-22
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-09-853-701-16

Query Match 44.3%; Score 39; DB 9; Length 281;
Best Local Similarity 52.6%; Pred. No. 88;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Oy 1 MKVSRVLAIVGAVIPAH 19
Db 164 VKLSOELVLPALIDFAHA 182

RESULT 12

US-09-925-300-1605
; Sequence 1605, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben

```

: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA101
: CURRENT APPLICATION NUMBER: US/09/925,300
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05988
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1890
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1605
: LENGTH: 326
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (30)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (31)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (116)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (182)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (226)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (285)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (287)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (290)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (298)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (306)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: US-09-925-300-1605

Query Match          44.3%; Score 39; DB 10; Length 326;
Best Local Similarity 56.2%; Pred. No. 1e+02;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY      4  SRVLALVIGAVIPAH 19
      1 1 1 1 1 1 1 1 1 1
Db      255  SNLALVLMGMOIPVHA 270

RESULT 13
US-09-815-242-14088
: Sequence 14088, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlssen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242

```

```

: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 14088
: LENGTH: 861
: TYPE: PRT
: ORGANISM: Salmonella typhi
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)...(861)
: OTHER INFORMATION: Xaa = Any Amino Acid
: US-09-815-242-14088

Query Match          44.3%; Score 39; DB 10; Length 861;
Best Local Similarity 53.3%; Pred. No. 3.1e+02;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY      1  MKVSRVLALVIGAVI 15
      1 1 1 1 1 1 1 1 1 1
Db      236  LKXRRVLALDMGALV 250

RESULT 14
US-09-925-299-1128
: Sequence 1128, Application US/09925299
: Publication No. US20030040617A9
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA102
: CURRENT APPLICATION NUMBER: US/09/925,299
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05983
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1556
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1128
: LENGTH: 130
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (116)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (122)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: US-09-925-299-1128

Query Match          43.2%; Score 38; DB 9; Length 130;
Best Local Similarity 44.4%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY      1  MKVSRVLALVIGAVIPAH 18
      1 1 1 1 1 1 1 1 1 1
Db      60  MKILRALNFGIGRPLPLH 77

```

```

RESULT 15
US-09-925-299-1128
: Sequence 1128, Application US/09925299
: Patent No. US20020055627A1
: GENERAL INFORMATION:
:   APPLICANT: Rosen et al.
:   TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
:   FILE REFERENCE: PA102
:   CURRENT APPLICATION NUMBER: US/09/925,299
:   CURRENT FILING DATE: 2001-08-10
:   PRIOR APPLICATION NUMBER: PCT/US00/05883
:   PRIOR FILING DATE: 2000-03-08
:   PRIOR APPLICATION NUMBER: 60/124,270
:   PRIOR FILING DATE: 1999-03-12
:   NUMBER OF SEQ. ID NOS: 1556
:   SOFTWARE: PatentIn Ver. 2.0
:   SEQ. ID NO 1128
:   LENGTH: 130
:   TYPE: PRT
:   ORGANISM: Homo sapiens
:   FEATURE:
:   NAME/KEY: SITE
:   LOCATION: (116)
:   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
:   NAME/KEY: SITE
:   LOCATION: (122)
:   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1128

Query Match          43.2%   Score 38; DB 10; Length 130;
Best Local Similarity 44.4%   Pred. No. 54;
Matches      8; Conservative      2; Mismatches      8; Indels      0; Gaps      0

OY      1 MKVSRVLVLGAVTPAH 18
        11::1111111111
DB      60 MKILRALNFGGRPLPLP 77

```

Search completed: April 26, 2003, 13:33:34
Job time : 0.749104 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: April 26, 2003, 13:33:03 ; Search time 1.27121 Seconds
(without alignments)
1436.866 Million cell updates/sec

Title: US-10-026-994-3

Perfect score: 88

Sequence: 1 MKVSRVLALVIGAVIPAH 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	52	59.1	339	2	H71361
2	46	52.3	333	2	I38974
3	46	52.3	486	2	H87311
4	44	50.0	95	2	A03380
5	44	50.0	397	2	A32370
6	44	50.0	483	2	A90176
7	43.5	49.4	1662	2	T18540
8	43.5	49.4	776	2	S13470
9	43.5	49.4	776	2	A37035
10	43	48.9	233	1	S13625
11	43	48.9	423	2	S34226
12	43	48.9	429	2	S34224
13	43	48.9	430	2	JH0509
14	43	48.9	430	2	A43285
15	43	48.9	430	2	I48316
16	43	48.9	485	2	T24677
17	43	48.9	540	2	T24675
18	43	48.9	896	2	S26984
19	43	48.9	2279	2	T42531
20	42	48.9	2280	2	T38906
21	42	47.7	365	2	A45260
22	42	47.7	395	2	D69779
23	42	47.7	483	2	JC7179
24	41.5	47.2	505	2	A63168
25	41	46.6	137	2	D72776
26	41	46.6	214	2	H81778
27	41	46.6	214	2	D81203
28	41	46.6	276	1	I39928
29	41	46.6	291	2	H84811

30	41	46.6	409	1	S32905	serine proteinase
31	41	46.6	492	2	A63539	cardiolipin synthe
32	41	46.6	773	2	I39597	hypothetical prote
33	41	46.6	955	2	T14361	H+-exporting ATPas
34	40	45.5	108	2	E70931	hypothetical prote
35	40	45.5	222	1	JL0085	eosinophil major b
36	40	45.5	234	1	S15102	eosinophil major b
37	40	45.5	258	2	S15955	entry exclusion pr
38	40	45.5	258	2	B46449	hypothetical prote
39	40	45.5	260	2	T35954	probable phosphati
40	40	45.5	315	2	D71012	hypothetical prote
41	40	45.5	322	2	JH0752	ribosomal protein
42	40	45.5	360	1	A43255	hydrogenase (EC 1.
43	40	45.5	360	1	S09250	hydrogenase (EC 1.
44	40	45.5	441	2	H83313	probable transport
45	40	45.5	527	2	T03427	hypothetical prote

ALIGNMENTS

RESULT 1

H71361

Conserved hypothetical integral membrane protein TP0151 - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999

C:Accession: H71361

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G

rson, J.J.; Khailak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M

they, L.; Weidman, J.; Smilt, R.O.; Venter, J.C.

A>Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770; PMID:9665876

A:Accession: H71361

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-339 <COL>

A:Cross-references: GB:AE001199; GB:AE000520; NID:g3322402; PIDN:AAC65135.1; PID:g332

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0151

Query Match

Best Local Similarity 59.1%; Score 52; DB 2; Length 339;

Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 SRVLALVIGAVIPAH 18

Db 74 SLVGLIGAVIPAH 88

RESULT 2

I38974

G protein-coupled receptor 8 - human

C:Species: Homo sapiens (man)

C>Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 21-Jul-2000

C:Accession: I38974

R:O'Dowd, B.F.; Scheidegger, M.A.; Nguyen, T.; Cheng, R.; Rasmussen, J.S.; Zastawny, R

Genomics 28, 84-91, 1995

A>Title: The cloning and chromosomal mapping of two novel human opioid-somatostatin-1

A:Reference number: A57647; MUID:96070436; PMID:7590751

A:Accession: I38974

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-333 <RES>

A:Cross-references: EMBL:U22492; NID:g953234; PIDN:AAC50198.1; PID:g953235

C:Genetics:

A:Gene: GDB:GPR8

A:Cross-references: GDB:371716; OMIM:600730

A:Map position: 20q13.3-20q13.3

C:Superfamily: vertebrate rhodopsin

C:Keywords: G protein-coupled receptor

Query Match	52.3%	Score 46	DB 2	Length 333
Best Local Similarity	66.7%	Pred. No. 11		
Matches 10	Conservative 1	Mismatches 4	Indels 0	Gaps 0
QY	2	KVSRKLYALYLCVAP	16	
		I I I I I I I I I I		
Db	210	KASRYVTLVLCGEVLP	224	

```

RESULT 3
H87311
hypothetical protein CC0505 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence.revision 20-Apr-2001 #text.change 20-Apr-2001
C:Accession: H87311
R:NIEMAN, W.C.; FELDBLYM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HELGEBERG, J.
B.; LABB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURLIN, A.S.; GWINN, M.L.; HAFT, D.H.; KOLCH
N. J.; ERMOLAeva, M.; WHITE, O.; SELZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87429; MUID:21173698; PMID:11259647
A:Accession: H87311
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <STO>
A:Cross-references: GB:AE005673; NID:q13421686; PIDN:AAK22492.1; GSPDB:GN00148
C:Genetics:
:Gene: CC0505

```

Query Match	52.3%	Score 46	DB 2	Length 486
Best Local Similarity	64.3%	Pred. No. 15		
Matches	9	Conservative	2	Mismatches 3
				Indels 0
				Gaps 0
QY	6	VLALVLCGAVTPAHA	19	
Db	11	VLAILGLGIAPPAHA	24	

```

RESULT 4
AD3380
outer membrane protein E BMEI1026 [imported] - Brucella melitensis (strain 16m)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AD3380
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patera, G.; Mijer, C.; Los, T.; Ivanova,
M.; Mazur, M.; Goldsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Lelessis,
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3380
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 195 <KUR>
A:Cross-references: GB:AE00917; PIDN:AAL52207.1; PID:g17982989; GSPDB:GN00190
A:Experimental source: strain 16m
C:Genetics:
A:Gene: BMEI1026
A:Map position: 1

```

Query Match	50.0%	Score 44	DB 2	Length 95
Best Local Similarity	42.1%	Pred. No. 7.4		
Matches	8	Conservative	6	Mismatches 5
				Indels 0
				Gaps 0
Oy	1	MKVSRYLALVCAVIPAHA	19	
		: : :		
Ob	1	MKIAGFAALLGSGTVMAHA	19	

RESULT 5
A32370
Cyclin B1 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
#:Date: 02-Nov-1989 #sequence_revision 02-Nov-1989
#text_change 16-Jul-1995

C:Accession: A32370
R:Minshull, J.; Blow, J.J.; Hunt, T.
Cell 56, 947-956, 1989
A:Title: Translation of cyclin mRNA is necessary for extracts of activated Xenopus egg
A:Reference number: A32370; MUID:89168446; PMID:2564315

A:Cross-references: GB:J03166; NID:g214092; PIDN:AAA49696.1; PID:g214093
C:Superfamily: cyclin
C:Keywords: cell cycle control

Query Match	50.0%	Score 44	DB 2	Length 397
Best Local Similarity	44.4%	Pred. No. 27		
Matches	8	Conservative	3	Mismatches
			7	Indels
			0	Gaps
QY	1	MKSVRYLALVIGAVIPAH	18	
		:	:	
Db	253	MKILRLEKFAIGRLPLH	270	

RESULT 6
A90176
NADH dehydrogenase subunit N (Nuon) Nuon [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: A90176
R:Shen, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Ch
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redden
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to Genbank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: A90176
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-483 <KUP>
A:Cross-references: GB:AE006641; NID:g13813474; PIDN:AAK40664.1; GSPDB:GN00155
A:Genetics:
A:Gene: Nuon

Query Match	50.0%	Score 44	DB 2	Length 483
Best Local Similarity	57.1%	Pred. No. 32		
Matches	8	Conservative	4	Mismatches
			2	Indels
			0	Gaps
QY	3	VSRLVALVGA	VIP 16	
		: ::		
db	459	VSALSIILGI	VIP 472	

```

RESULT 7
T18540
mofa protein.precursor [imported] - Leptochrix discophora
C:Species: Leptochrix discophora
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 08-Sep-2000
C:Accession: T18540
R:Corstjens, P.L.
submitted to the EMBL Data Library, April 1999
A:Reference number: Z18959
A:Accession: T18540
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1662 <COR>
A:Cross-references: EMBL:Z25774; NID:e1427784; PTD:e267517; PIDN:CAAB1037.1.
C:Genetics:
A:Note: mofa
F:1-33/Domain: signal sequence #status predicted <SIG>
F:43-1662/Product: mofa protein #status predicted <MAT>

```

Query Match	50.0%;	Score 44;	DB 2;	Length 1662;
Best Local Similarity	64.3%;	Pred. No. 95;		
Matches	9;	Conservative	3;	Mismatches 2;
				Indels 0;
				Gaps 0;

QY 6 VLATVLCGAVIPAH 19
:||||:||||
Db 18 LLALLLGAAMPAPA 31

RESULT 8

isoamylase (EC 3.2.1.68) precursor - Pseudomonas sp.

C:Species: Pseudomonas sp.

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 15-Oct-1999

C:Accession: S13470; A28109; A45919

R:Chen, J.H.; Chen, Z.Y.; Chow, T.Y.; Chen, J.C.; Tan, S.T.; Hsu, W.H.

C:Biochim. Biophys. Acta 1087, 309-315, 1990

A:Title: Nucleotide sequence and expression of the isoamylase gene from an isoamylase-hy

A:Reference number: S13470; MUID:91064385; PMID:2248978

A:Accession: S13470

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-776 <CHE>

R:Amemura, A.; Chakraborty, R.; Fujita, M.; Nouni, T.; Futai, M.

J. Biol. Chem. 263, 9271-9275, 1988

A:Title: Cloning and nucleotide sequence of the isoamylase gene from Pseudomonas amylo

A:Reference number: A28109; MUID:88243808; PMID:3379068

A:Accession: A28109

A:Molecule type: DNA

A:Residues: 1-776, 'G', '9-125', 'C', '127-168', '170', 'H', '172-385', 'V', '387-412', 'AVH', '417-453', 'T', 'V',

A:Note: the source is designated as Pseudomonas amyloclavum

R:Fujita, M.; Amemura, A.; Futai, M.

J. Bacteriol. 171, 4320-4323, 1989

A:Title: Transcription of the isoamylase gene (iam) in Pseudomonas amyloclavum SB-15.

A:Reference number: A45919; MUID:89327147; PMID:2753857

A:Accession: A45919

A:Molecule type: DNA

A:Residues: 1-776, 'G', '9-25' <FNU>

A:Cross-references: GB:M28370

C:Superfamily: glyx protein

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-26/Domain: signal sequence #status predicted <SIG>

F:27-776/Product: isoamylase #status experimental <MAT>

Query Match 49.4%; Score 43.5; DB 2; Length 776;
Best Local Similarity 38.5%; Pred. No. 58;
Matches 10; Conservative 4; Mismatches 5; Indels 7; Gaps 1;

QY 1 MKVSRVLCGAVI-----PAHA 19
||:||||:||||
Db 1 MKCPRLALLGCAVLGVPMPAPA 26

RESULT 9

isoamylase (EC 3.2.1.68) precursor - Pseudomonas sp. (strain SMP1)

C:Species: Pseudomonas sp.

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 26-Aug-1999

C:Accession: A37035; B37035

R:Tognoni, A.; Carreira, P.; Galli, G.; Lucchese, G.; Camerini, B.; Grandi, G.

J. Gen. Microbiol. 135, 37-45, 1989

A:Title: Cloning and nucleotide sequence of the isoamylase gene from a strain of Pseudom

A:Reference number: A37035; MUID:89381677; PMID:2778432

A:Accession: A37035

A:Molecule type: DNA

A:Residues: 1-776 <TOG>

A:Cross-references: GB:M25247; NID:g151295; PIDN:AAA25855.1; PID:g151296

A:Accession: B37035

A:Molecule type: protein

A:Residues: 27-44 <TO2>

C:Comment: Isoamylases hydrolyze alpha-1,6 glucosidic branch linkages in glycogen and am

C:Superfamily: glyx protein

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

F:1-26/Domain: signal sequence #status predicted <SIG>

F:27-776/Product: isoamylase #status predicted <MAT>

Query Match 49.4%; Score 43.5; DB 2; Length 776;
Best Local Similarity 38.5%; Pred. No. 58;
Matches 10; Conservative 4; Mismatches 5; Indels 7; Gaps 1;

QY 1 MKVSRVLCGAVI-----PAHA 19
||:||||:||||
Db 1 MKCPRLALLGCAVLGVPMPAPA 26

RESULT 10

eosinophil major basic protein 1 precursor - guinea pig

C:Species: Cavia porcellus (guinea pig)

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 21-Jul-2000

C:Accession: S13625; S16497; S40491

R:Aoki, I.; Shindoh, Y.; Nishida, T.; Nakai, S.; Hong, Y.M.; Mio, M.; Saito, T.; Tasa

FEBS Lett. 279, 330-334, 1991

A:Title: Sequencing and cloning of the cDNA of guinea pig eosinophil major basic prot

A:Reference number: S13625; MUID:91160746; PMID:1705901

A:Accession: S13625

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-233 <ACK>

A:Cross-references: GB:D90251; NID:g3135096; PIDN:BA14291.1; PID:g220291

A:Accession: S18497

A:Molecule type: protein

A:Residues: 115-146, 'X', '148-157', 'X', '159-162' <AOK1>

R:Hashimoto, Y.; Nagaoaka, I.; Yamashita, T.

Biophys. Acta 1203, 236-242, 1993

A:Title: Purification of the antibacterial fragments of guinea-pig major basic protei

A:Reference number: S40491; MUID:94092714; PMID:8268206

A:Accession: S40491

A:Molecule type: protein

A:Residues: 115-233 <HAS>

C:Superfamily: eosinophil major basic protein precursor; C-type lectin homology

C:Keywords: antibiotic; chondroitin sulfate proteoglycan; cytotoxin; eosinophil; glyx

F:17-114/Domain: signal sequence #status predicted <SIG>

F:115-233/Product: eosinophil major basic protein 1 #status experimental <MAT>

F:116-231/Domain: C-type lectin homology <LCH>

F:24-25/Binding site: carbohydrate (Ser) (covalent) #status predicted

F:67/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

F:134-231,208-223/Disulfide bonds: #status predicted

Query Match 48.9%; Score 43; DB 1; Length 233;
Best Local Similarity 55.6%; Pred. No. 24;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MKVSRVLCGAVIPAH 18
||:||||:||||
Db 1 MKLLALLLGLGAVSTRH 18

RESULT 11

cyclin B - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 21-Jul-2000

C:Accession: S34226; S20658; I52833

R:Markiewicz, D.A.; Flick, M.B.; Mushel, R.J.; McKenna, W.G.

submitted to the EMBL data library, March 1992

A:Description: New features of mammalian cyclins seen in rat and chinese hamster cycl

A:Reference number: S34224

A:Accession: S34226

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-423 <MAR>

A:Cross-references: EMBL:X64589; NID:g313807; PIDN:CAA45877.1; PID:g313808

R:Kenacka, Y.; Nojima, H.; Okayama, H.

submitted to the EMBL data library, July 1991

A:Description: Nucleotide sequences of cDNAs encoding rat cdc2 + and cyclin 2.

A:Accession: S20658
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-423 <KAN>
A:Cross-references: EMBL:X60768; NID:g56027; PIDN:CAA43178.1; PID:g56028
R:Trembley, J.H.; Kren, B.T.; Steer, C.J.
Cell Growth Differ. 5, 99-108, 1994
A:Title: Posttranscriptional regulation of cyclin B messenger RNA expression in the reg
A:Reference number: 152833; MUID:94169087; PMID:8123599
A:Accession: 152833
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-423 <RES>
A:Cross-references: GB:L11995; NID:g2828516; PIDN:AAC00032.1; PID:g203706
C:Superfamily: cyclin
C:Keywords: cell cycle control

Query Match 48.9%; Score 43; DB 2; Length 423;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 MKVSRVLALVIGAVIPAH 18
DB 277 MKILRVLFSLGRPLPLH 294

RESULT 12
S34224
cyclin B - long-tailed hamster
C:Species: Cricetus longicaudatus (long-tailed hamster)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jul-1999
C:Accession: S34224
R:Markiewicz, D.A.; Flick, M.B.; Mushel, R.J.; McKenna, W.G.
submitted to the EMBL Data Library, March 1992
A:Description: New features of mammalian cyclins seen in rat and chinese hamster cyclin
A:Reference number: S34224
A:Accession: S34224
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-429 <MAR>
A:Cross-references: EMBL:X64588; NID:g313764; PIDN:CAA45876.1; PID:g313765
C:Superfamily: cyclin
C:Keywords: cell cycle control

Query Match 48.9%; Score 43; DB 2; Length 429;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 MKVSRVLALVIGAVIPAH 18
DB 283 MKILRVLFSLGRPLPLH 300

RESULT 13
JH0509
cyclin B - mouse
C:Species: Mus musculus (house mouse)
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999
C:Accession: JH0509; S14652
R:Paterno, G.D.; Downs, K.M.
Gene 108, 315-316, 1991
A:Title: Sequence of a cDNA encoding a mouse cyclin B protein.
A:Reference number: JH0509; MUID:92084156; PMID:1836195
A:Accession: JH0509
A:Molecule type: mRNA
A:Residues: 1-430 <PAR>
A:Cross-references: EMBL:X58708; NID:g50614; PIDN:CAA41545.1; PID:g50615
A:Experimental source: embryonic carcinoma cell
C:Comment: The cyclins are crucial components of a protein kinase cascade regulating a c
C:Superfamily: cyclin
C:Keywords: cell cycle control

Query Match 48.9%; Score 43; DB 2; Length 430;

Best Local Similarity 50.0%; Pred. No. 41;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 MKVSRVLALVIGAVIPAH 18
DB 284 MKILRVLFSLGRPLPLH 301

RESULT 14
A43285
cyclin B1 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
C:Accession: A43285
R:Hanley-Hyde, J.; Mushinski, J.F.; Sadofsky, M.; Huppl, K.; Krall, M.; Kozak, C.A.;
Genomics 13, 1018-1030, 1992
A:Title: Expression of murine cyclin B1 mRNAs and genetic mapping of related genomic
A:Reference number: A43285; MUID:92371993; PMID:1387105
A:Accession: A43285
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-430 <HAN>
A:Cross-references: GB:S43105; NID:g254021; PIDN:AAB22970.1; PID:g254022
A:Note: sequence extracted from NCBI backbone (NCBIN:111726, NCBIPI:111727)
C:Superfamily: cyclin

Query Match 48.9%; Score 43; DB 2; Length 430;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 MKVSRVLALVIGAVIPAH 18
DB 284 MKILRVLFSLGRPLPLH 301

RESULT 15
I48316
cyclin B1 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Jul-1999
C:Accession: I48316
R:Chapman, D.L.; Wolgemuth, D.J.
Mol. Reprod. Dev. 33, 259-269, 1992
A:Title: Identification of a mouse B-type cyclin which exhibits developmentally regul
A:Reference number: I48316; MUID:93080989; PMID:1280449
A:Accession: I48316
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-430 <RES>
A:Cross-references: EMBL:X64713; NID:g50616; PIDN:CAA45968.1; PID:g50617
C:Genetics:
A:Gene: cyclB1
C:Superfamily: cyclin

Query Match 48.9%; Score 43; DB 2; Length 430;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 MKVSRVLALVIGAVIPAH 18
DB 284 MKILRVLFSLGRPLPLH 301

Search completed: April 26, 2003, 13:38:30
Job time : 1.27121 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:26:23 ; Search time 0.726404 Seconds
(without alignments)
1084.865 Million cell updates/sec

Title: US-10-026-994-3
Perfect score: 88
Sequence: 1 MKVSRVLALVIGAVIPARA 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	52.3	333	1 GPR8_HUMAN	P48146 homo sapien
2	44	50.0	397	1 GGB1_XENLA	P13350 xenopus lae
3	43.5	49.4	776	1 ISOA_PSEAY	P10342 pseudomonas
4	43.5	49.4	776	1 ISOA_PSEAY	P26501 pseudomonas
5	43	48.9	233	1 EMB1_CAVPO	P22032 cavia porce
6	43	48.9	423	1 GGB1_CAVPO	P30277 ratulus norv
7	43	48.9	429	1 GGB1_CAVPO	P08301 cricetus
8	43	48.9	429	1 GGB1_CAVPO	P37882 mesocricetu
9	43	48.9	430	1 GGB1_CAVPO	P24880 mus musculu
10	43	48.9	896	1 RPOP_NEUCR	P33540 neurospora
11	43	48.9	2280	1 COAC_SCHPO	P78820 schizosacch
12	42	47.7	365	1 SHLE_HUMAN	P28566 homo sapien
13	41	46.6	276	1 SSUC_BACSU	P40401 bacillus su
14	41	46.6	409	1 ALP_TRIHA	P03420 trichoderma
15	41	46.6	773	1 HEBX_ALTSO	P49007 altermonas
16	40	45.5	222	1 EMB1_HUMAN	P13727 homo sapien
17	40	45.5	234	1 EMB1_CAVPO	P35709 cavia porce
18	40	45.5	323	1 RLAO_TRYCR	P26796 trypanosoma
19	40	45.5	360	1 MBHS_ALCEU	P31892 alcaligenes
20	40	45.5	360	1 MBHS_ALCEU	P13633 rhodocyclu
21	40	45.5	574	1 GLYA_TREPA	P08339 treponema p
22	40	45.5	618	1 CIRI_CITFR	P23182 citrobacter
23	39.5	44.9	403	1 ORYZ_ASPFU	P28296 aspergillus
24	39.5	44.9	476	1 RMUC_SALTU	P01567 salmonella
25	39	44.3	350	1 PORF_PSEAE	P13794 pseudomonas
26	39	44.3	463	1 MFGM_MOUSE	P21956 mus musculu
27	39	44.3	616	1 SYGM_YEAST	P22438 saccharomyc
28	39	44.3	616	1 YGO2_YEAST	P30777 saccharomyc
29	39	44.3	677	1 FLHA_BACSU	P35620 bacillus su
30	39	44.3	712	1 CDGT_BACSU	P09121 bacillus su
31	38.5	43.8	713	1 CDGT_BACSU	P05618 bacillus sp
32	38.5	43.8	236	1 APHA_HAEIN	P44009 haemophilus
33	38	43.2	146	1 GPR7_MOUSE	P49681 mus musculu

34	38	43.2	218	1 CIB1_HUMAN	Q07699 homo sapien
35	38	43.2	252	1 MODA_RHOCA	Q08383 rhodobacter
36	38	43.2	303	1 MTRF_METJA	Q58257 methanococc
37	38	43.2	304	1 MTRF_METJA	P80551 methanosarc
38	38	43.2	322	1 RBSC_BACSU	P36948 bacillus su
39	38	43.2	328	1 GPR7_HUMAN	P48145 homo sapien
40	38	43.2	379	1 CYB_PANLE	Q35506 panthera le
41	38	43.2	414	1 MTR_ECOLI	Q35655 panthera ti
42	38	43.2	414	1 MTR_ECOLI	P22306 escherichia
43	38	43.2	434	1 GPR7_HUMAN	P14635 homo sapien
44	38	43.2	434	1 CBPS_STRCP	P39041 streptomyce
45	38	43.2	485	1 ER24_FUSSO	Q01447 fusarium so

ALIGNMENTS

RESULT 1
ID GPR8_HUMAN STANDARD: PRT; 333 AA.
AC P48146: Q9H4K3:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable G protein-coupled receptor GPR8.
GN GPR8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID:9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96070436; PubMed=7590751;
RA O'Dowd B.F., Scheideler M.A., Nguyen T., Cheng R., Rasmussen J.S.,
RA Marchese A., Zaslansky R., Heng H.H.Q., Tsui L.-C., Shi X., Asa S.,
RA Puy L., George S.R.;
RT "The cloning and chromosomal mapping of two novel human opiod-
RT somatostatin-like receptor genes, GPR7 and GPR8, expressed in
RT discrete areas of the brain.";
RL Genomics 28:84-91(1995).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.V., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grahnam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hackie E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasistaho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McCormachie L.J., McElay K., McNurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Strue C.D., Smith M.L., Soderlund C., Steward G.K., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Walli M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilmshurst L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:665-871(2001).
[3]
RP VARIANT GLN-206.
RA Zeng J., Liu M., Grau O., Capron A., Bahr G.M.;

"Identification of a novel amino acid substitution (R206Q) in the second extracellular loop of the opioid-somatostatin-like receptor RT gene GPR8.".

RT Hum. Mutat. 12:219-219(1998).

CC -1- FUNCTION: OPIAN RECEPTOR. COULD BIND AN OPIOID.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- TISSUE SPECIFICITY: FRONTAL CORTEX. NOT IN CEREBELLUM, HIPPOCAMPUS, HYPOTHALAMUS, PONS, PUTAMEN AND THALAMUS REGIONS.

CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U22492; AAC50198.1; -.

CC EMBL; AL121581; CAC17004.1; -.

CC Genew; HGNC:4530; GPR8.

CC MIM; 600731; -.

CC Interpro: IPR000276; GPCR_Rhodopsn.

CC Pfam; PR00237; GPCR_Rhodopsn.

CC PRINTS; PR00237; GPCR_Rhodopsn.

CC PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.

CC PROSITE; PS00237; G-PROTEIN_RECEP_F1_2; 1.

CC G-Protein coupled receptor; Transmembrane; Glycoprotein; Lipoprotein; K1 Palmitate; Phosphorylation; Polymorphism.

CC DOMAIN 1 45 EXTRACELLULAR (POTENTIAL).

CC FT TRANSSEM 46 69 1 (POTENTIAL).

CC FT DOMAIN 70 80 CYTOPLASMIC (POTENTIAL).

CC FT TRANSSEM 81 105 2 (POTENTIAL).

CC FT DOMAIN 106 120 EXTRACELLULAR (POTENTIAL).

CC FT TRANSSEM 121 140 3 (POTENTIAL).

CC FT DOMAIN 141 165 CYTOPLASMIC (POTENTIAL).

CC FT TRANSSEM 166 185 4 (POTENTIAL).

CC FT DOMAIN 186 211 EXTRACELLULAR (POTENTIAL).

CC FT TRANSSEM 212 233 5 (POTENTIAL).

CC FT DOMAIN 234 257 CYTOPLASMIC (POTENTIAL).

CC FT TRANSSEM 258 282 6 (POTENTIAL).

CC FT DOMAIN 283 292 EXTRACELLULAR (POTENTIAL).

CC FT TRANSSEM 293 307 7 (POTENTIAL).

CC FT DOMAIN 308 333 CYTOPLASMIC (POTENTIAL).

CC FT DISULFID 117 197 BY SIMILARITY.

CC FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 29 29 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC FT VARIANT 206 206 R -> Q.

CC FT FTID=VAR_003579.

CC FT CONFLICT 305 305 T -> S (IN REF. 2).

CC FT SEQUENCE 333 AA; 36902 MW; 2343B473FAF217BB CRC64;

Query Match 52.38; Score 46; DB 1; Length 333;
Best Local Similarity 66.78; Pred. No. 6.3;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 KVSRLVLCGAVIP 16
1 111 111 111 111 111
Db 210 KASRVYTLVGLFVLP 224

RESULT 2
CGBI_XENLA STANDARD; PRT; 397 AA.
AC P13350;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE G2/mitotic-specific cyclin B1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89168446; PubMed=2564315;
RA Minshull J., Blow J.J., Hunt T.;
RT "Translation of cyclin mRNA is necessary for extracts of activated
RT xenopus eggs to enter mitosis.";
RL Cell 56:947-956(1989).

CC -1- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
CC (MITOSIS) TRANSITION.

CC -1- SUBUNIT: INTERACTS WITH THE CDC2 PROTEIN KINASE TO FORM A
CC SERINE/THREONINE KINASE HOLOENZYME COMPLEX ALSO KNOWN AS
CC NUCLEATION PROMOTING FACTOR (NPF). THE CYCLIN SUBUNIT IMPARTS
CC SUBSTRATE SPECIFICITY TO THE COMPLEX.

CC -1- DEVELOPMENTAL STAGE: ACCUMULATES STEADILY DURING G2 AND IS
CC ABRUPTLY DESTROYED AT MITOSIS.

CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; J03166; AAA49696.1; -.

CC PIR; A32370; A32370.

CC HSSP; P30274; IVIN.

CC Interpro: IPR004366; Cyclin.

CC Interpro: IPR004367; Cyclin_Cterm.

CC Pfam; PF00134; cyclin.1.

CC Pfam; PF02984; cyclin.C; 1.

CC SMART; SM00385; CYCLIN; 2.

CC PROSITE; PS00292; CYCLINS; 1.

CC CYCLIN; Cell cycle; Cell division; Mitosis.

CC FT DOMAIN 45 77 LYS-RICH.

CC FT SEQUENCE 397 AA; 44673 MW; FBDA648E1258D78F CRC64;

Query Match 50.08; Score 44; DB 1; Length 397;
Best Local Similarity 44.48; Pred. No. 15;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 MKVSRLVLCGAVIPAH 18
1 111 111 111 111 111
Db 253 MKILRVLCFALGRPLPH 270

RESULT 3
ISOA_PSEAY STANDARD; PRT; 776 AA.
ID ISOA_PSEAY
AC P10342;
DT 01-MAR-1989 (Rel. 10, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Isoamylase precursor (EC 3.2.1.68).
GN IAM.
OS Pseudomonas amylofermosa.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=32043;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB-15;
RX MEDLINE=88243808; PubMed=3379068;
RA Amemura A., Chakraborty R., Fujita M., Nouni T., Futai M.;
RT "Cloning and nucleotide sequence of the isoamylase gene from
RT Pseudomonas amylofermosa SB-15.";
RL J. Biol. Chem. 263:9271-9275(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JD210;
RX MEDLINE=91064385; PubMed=2248978;

RA Chen J.H., Chen Z.Y., Chow T.Y., Chen J.C., Tan S.T., Hsu W.H.;
 RT "Nucleotide sequence and expression of the isoamylase gene from an
 RT isoamylase-hyperproducing mutant, *Pseudomonas amyloclavata* JD210.";
 RL Biochim. Biophys. Acta 1087:309-315(1990).
 RN [3]
 RP SEQUENCE OF 744-776 FROM N.A.
 RC STRAIN-SB-15;
 RX MEDLINE=89327147; PubMed=2753857;
 RA Amemura A., Fujita M., Futai M.;
 RT "Transcription of the isoamylase gene (iam) in *Pseudomonas*
 RT *amyloclavata* SB-15.";
 RN J. Bacteriol. 171:4320-4325(1989).
 RL [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=98387895; PubMed=9719642;
 RA Katsuya Y., Mezaki Y., Kubota M., Matsuura Y.;
 RT "Three-dimensional structure of *Pseudomonas* isoamylase at 2.2-A
 RT resolution.";
 RL J. Mol. Biol. 281:885-897(1998).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(1,6)-D-glucosidic branch
 CC linkages in glycogen, amylopectin and their beta-limits dextrins.
 CC -1- INDUCTION: BY MALTOS. amylopectin and their beta-limits dextrins.
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY. ISOAMYLASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J03871; AAA25854.1; -;
 DR EMBL: X13378; CAA31754.1; -;
 DR PIR: A28109; A28109.
 DR PIR: 1BF2; 12-AUG-98.
 DR InterPro: IPR000461; Alpha-amylase.
 DR InterPro: IPR004193; Isoamylase.N.
 DR Pfam: PF00128; alpha-amylase.N.
 DR Pfam: PF02922; isoamylase.N; 1.
 DR Pfam: PF02922; isoamylase.N; 1.
 KW Hydrolyase; Glycosidase; Signal; 3D-structure.
 FT SIGNAL 1 26
 FT CHAIN 27 776
 FT ACT_SITE 401 401 ISOAMYLASE.
 FT ACT_SITE 481 481 BY SIMILARITY.
 FT ACT_SITE 536 536 BY SIMILARITY.
 FT DISULFID 410 422 BY SIMILARITY.
 FT DISULFID 546 616 BY SIMILARITY.
 FT DISULFID 738 766 BY SIMILARITY.
 FT DISULFID 738 766 BY SIMILARITY.
 FT CONFLICT 8 8
 FT CONFLICT 126 126 A -> G (IN REF. 1).
 FT CONFLICT 169 171 F -> C (IN REF. 1).
 FT CONFLICT 386 386 GAS -> AH (IN REF. 1).
 FT CONFLICT 413 416 L -> V (IN REF. 1).
 FT CONFLICT 454 489 GAYT -> AVH (IN REF. 1).
 FT CONFLICT 454 489 SGLDFAEPNAGISYOLGEPQSGMSEWNLGFRDS ->
 FT CONFLICT 454 489 TWICLRLRLGPSAATRTSWSVRVREMSVPRQ (IN
 FT CONFLICT 454 489 REF. 1).
 FT CONFLICT 555 556 WP -> S (IN REF. 1).
 FT CONFLICT 650 657 AFKRAHPA -> RSRARHP (IN REF. 1).
 FT CONFLICT 657 657 AFKRAHPA -> RSRARHP (IN REF. 1).
 FT SEQUENCE 776 AA; 83626 MW; F738BF8040246169 CRC64;

Query Match 49.4%; Score 43.5; DB 1; Length 776;
 Best Local Similarity 38.5%; Pred. No. 30;
 Matches 10; Conservative 4; Mismatches 5; Indels 7; Gaps 1;

OY 1 MKVSRVLAIVGAVI-----PAHA 19
 DB 1 MKCPKILALIGCAVLAIVGAVPAMPAPA 26

RESULT 4
 ISOA_PSESP

ID ISOA_PSESP STANDARD; PRT: 776 AA.
 AC P25501;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Isoamylase precursor (EC 3.2.1.68).
 GN IAM.
 OS *Pseudomonas* sp. (strain SMP1).
 CC Bacteria; Proteobacteria.
 CC NCB1_TaxID=306;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-44.
 RX MEDLINE=89381677; PubMed=2778432;
 RA Tognoni A., Carrera P., Galli G., Lucchese G., Camerini B.,
 RA Grandi G.;
 RT "Cloning and nucleotide sequence of the isoamylase gene from a strain
 RT of *Pseudomonas* sp.";
 RL J. Gen. Microbiol. 135:37-45(1989).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(1,6)-D-glucosidic branch
 CC linkages in glycogen, amylopectin and their beta-limits dextrins.
 CC -1- INDUCTION: BY MALTOS. amylopectin and their beta-limits dextrins.
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY. ISOAMYLASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M25247; AAA25855.1; -;
 DR EMBL: A10909; CAA00929.1; -;
 DR PIR: A37035; A37035.
 DR HSSP: P10342; 1BF2.
 DR InterPro: IPR000461; Alpha-amylase.
 DR InterPro: IPR004193; Isoamylase.N.
 DR Pfam: PF00128; alpha-amylase.N.
 DR Pfam: PF02922; isoamylase.N; 1.
 DR Pfam: PF02922; isoamylase.N; 1.
 KW Hydrolyase; Glycosidase; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 776
 FT ACT_SITE 401 401 ISOAMYLASE.
 FT ACT_SITE 481 481 BY SIMILARITY.
 FT ACT_SITE 536 536 BY SIMILARITY.
 FT DISULFID 410 422 BY SIMILARITY.
 FT DISULFID 546 616 BY SIMILARITY.
 FT DISULFID 738 766 BY SIMILARITY.
 FT DISULFID 738 766 BY SIMILARITY.
 FT SEQUENCE 776 AA; 83656 MW; ASE4C02EF026A3A4 CRC64;

Query Match 49.4%; Score 43.5; DB 1; Length 776;
 Best Local Similarity 38.5%; Pred. No. 30;
 Matches 10; Conservative 4; Mismatches 5; Indels 7; Gaps 1;

OY 1 MKVSRVLAIVGAVI-----PAHA 19
 DB 1 MKCPKILALIGCAVLAIVGAVPAMPAPA 26

RESULT 5
 EMBL_CAVPO
 ID EMBL_CAVPO STANDARD; PRT: 233 AA.
 AC P22032;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Eosinophil granule major basic protein 1 precursor (MBP-1).
 GN MBP.
 OS *Cavia porcellus* (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriocognath; Caviidae; Cavia.
 OC NCB1_TaxID=10141;

RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 115-162.
 RC TISSUE-Eosinophil;
 RX MEDLINE-91160746; PubMed-1705901;
 RA Aoki I., Shindoh Y., Nishida T., Nakai S., Hong Y.-M., Mio M.,
 RT Saito T., Tasaka K.;
 RT Sequencing and cloning of the cDNA of guinea pig eosinophil major
 RT basic protein.";
 RL FEBS Lett. 279:330-334(1991).
 RN [2]
 RP PARTIAL SEQUENCE.
 RX MEDLINE-94092714; PubMed-8268206;
 RA Hashimoto Y., Nagaoaka I., Yamashita T.;
 RT "Purification of the antibacterial fragments of guinea-pig major
 RT basic protein.";
 CC -1- FUNCTION: MBP MAY PLAY SOME IMPORTANT ROLES IN THE ALLERGIC
 CC REACTIONS AND INFLAMMATIONS, SINCE MBP IS CAPABLE OF RELEASING
 CC HISTAMINE FROM MAST CELLS AND DAMAGING THE EPITHELIAL CELLS OF
 CC BRONCHIAL TUBES. ANTIPARASITIC AND ANTIBIOTIC.
 CC -1- SUBCELLULAR LOCATION: MATRIX OF EOSINOPHIL'S LARGE SPECIFIC
 CC GRANULE (CRYSTALLOID CORE).
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D90251; BAA14291.1; -
 DR PIR: S13625; S13625.
 DR HSSP: P13727; 1480.
 DR InterPro: IPR002352; EmaJor.bas.c.
 DR InterPro: IPR001304; Lectin.C.
 DR Pfam: PF00059; Lectin.C.1.
 DR PRINTS: PR00770; EMAJORBASICP.
 DR SMART: SM00034; CLECT.1.
 DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
 DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
 DR Eosinophil; Toxin; Signal; Cytotoxin; Immune response; Antibiotic;
 KW Lectin; Multigene family.
 FT SIGNAL 1 15 POTENTIAL.
 FT PROPEP 16 114 ACIDIC.
 FT CHAIN 115 233 EOSINOPHIL GRANULE MAJOR BASIC PROTEIN 1.
 FT DOMAIN 132 233 C-TYPE LECTIN (SHORT FORM).
 FT DISULFID 134 231 BY SIMILARITY.
 FT DISULFID 208 223 BY SIMILARITY.
 SQ SEQUENCE 233 AA; 26268 MW; C8D5E96D927C56C8 CRC64;
 Query Match 48.9%; Score 43; DB 1; Length 233;
 Best Local Similarity 55.6%; Pred. No. 14;
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MKVSRVLAIVLCAVIPA 18
 DB 1 MKLLLLALLLGVSTRH 18
 RESULT 6
 CGBI_RAT STANDARD: PRT: 423 AA.
 AC P30277;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE G2/mitotic-specific cyclin B1.
 GN CCNBI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID-10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kanoka Y., Nojima H., Okayama H.;
 RT Submitted (XXX-1992) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE-Fibroblast;
 RX MEDLINE-94095439; PubMed-8270434;
 RA Markiewicz D.A., McKenna W.G., Flick M.B., Maity A., Muschel R.J.;
 RT "The effects of radiation on the expression of a newly cloned and
 RT characterized rat cyclin B mRNA.";
 RL Int. J. Radiat. Oncol. Biol. Phys. 28:135-144(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE-Testis;
 RX MEDLINE-94169087; PubMed-8123599;
 RA Trembley J.H., Kren B.T., Steer C.J.;
 RT "Posttranscriptional regulation of cyclin B messenger RNA expression
 RT in the regenerating rat liver.";
 RL Cell Growth Differ. 5:99-108(1994).
 CC -1- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
 CC (MITOSIS) TRANSITION.
 CC -1- SUBUNIT: INTERACTS WITH THE CDC2 PROTEIN KINASE TO FORM A
 CC SERINE/THREONINE KINASE HOLOENZYME COMPLEX ALSO KNOWN AS
 CC MATURATION PROMOTING FACTOR (MPF). THE CYCLIN SUBUNIT IMPARTS
 CC SUBSTRATE SPECIFICITY TO THE COMPLEX.
 CC -1- DEVELOPMENTAL STAGE: ACCUMULATES STEADILY DURING G2 AND IS
 CC ABRUPTLY DESTROYED AT MITOSIS.
 CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X60768; CAA43178.1; -
 DR EMBL: X64589; CAA45877.1; -
 DR EMBL: L11995; AAC00032.1; -
 DR PIR: S20658; S20658.
 DR PIR: S34226; S34226.
 DR HSSP: P30274; 1VIN.
 DR InterPro: IPR004367; Cyclin.
 DR InterPro: IPR004367; Cyclin.cterm.
 DR Pfam: PF00134; cyclin; 1.
 DR Pfam: PF02984; cyclin.C; 1.
 DR SMART: SM00385; CYCLIN; 2.
 DR PROSITE: PS00292; CYCLINS; 1.
 KW Cyclin; Cell cycle; Cell division; Mitosis.
 FT DOMAIN 50 82 LYS-RICH.
 SQ SEQUENCE 423 AA; 47391 MW; 4223CF71F144B279 CRC64;
 Query Match 48.9%; Score 43; DB 1; Length 423;
 Best Local Similarity 50.0%; Pred. No. 22;
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 1 MKVSRVLAIVLCAVIPA 18
 DB 277 MKILRVNLNFSRLPLH 294
 RESULT 7
 CGBI_CRITO STANDARD: PRT: 429 AA.
 AC Q08301;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE G2/mitotic-specific cyclin B1.
 GN CCNBI.

```

OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10030;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=94095439; PubMed=8270434;
RA Martlewicz D.A., McKenna W.G., Flick M.B., Maity A., Muschel R.J.;
RT "The effects of irradiation on the expression of a newly cloned and
RT characterized rat cyclin B mRNA."
RT Int. J. Radiat. Oncol. Biol. Phys. 28:135-144(1994).
CC CC
CC -1- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
CC (MITOSIS) TRANSITION.
CC -1- SUBUNIT: INTERACTS WITH THE CDC2 PROTEIN KINASE TO FORM A
CC SERINE/THREONINE KINASE HOLOENZYME COMPLEX ALSO KNOWN AS
CC MATURATION PROMOTING FACTOR (MPF). THE CYCLIN SUBUNIT IMPARTS
CC SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC -1- DEVELOPMENTAL STAGE: ACCUMULATES STEADILY DURING G2 AND IS
CC ABRUPTLY DESTROYED AT MITOSIS.
CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: X64588; CAA45876.1; -
DR PIR: S34224; S34224.
DR HSSP: P30274; IVIN.
DR InterPro: IPR004367; Cyclin_Cterm.
DR InterPro: IPR004367; Cyclin_Cterm.
DR Pfam: PF00134; cyclin; 1.
DR Pfam: PF02984; cyclin_C; 1.
DR SMART: SM00385; CYCLIN; 2.
DR PROSITE: PS00292; CYCLINS; 1.
DR Cyclin; Cell cycle; Cell division; Mitosis.
KW DOMAIN
FT 51 86 LYS-RICH.
SQ SEQUENCE: 429 AA; 48062 MW; 6E0BAE751A678B7 CRC64;

Query Match 48.9%; Score 43; DB 1; Length 429;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 MKVSRVLAIVLGVIPAH 18
DB 283 MKILRVNFSLGRLPLH 300
II: ||| | | |
|: ||| | | |

RESULT 8
CG1L_MESAU STANDARD; PRT; 429 AA.
AC P37882;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE G2/mitotic-specific cyclin B1.
GN CCNB1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Shiraki T., Yamashita K., Nishitani H., Nishimoto T.;
RT Submitted (Jul-1993) to the EMBL/GenBank/DBI databases.
CC -1- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
CC (MITOSIS) TRANSITION.

```

```

CC -1- SUBUNIT: INTERACTS WITH THE CDC2 PROTEIN KINASE TO FORM A
CC SERINE/THREONINE KINASE HOLOENZYME COMPLEX ALSO KNOWN AS
CC MATURATION PROMOTING FACTOR (MPF). THE CYCLIN SUBUNIT IMPARTS
CC SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC -1- DEVELOPMENTAL STAGE: ACCUMULATES STEADILY DURING G2 AND IS
CC ABRUPTLY DESTROYED AT MITOSIS.
CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: D17293; BA04126.1; -
DR HSSP: P30274; IVIN.
DR InterPro: IPR004366; Cyclin.
DR InterPro: IPR004367; Cyclin_Cterm.
DR Pfam: PF00134; cyclin; 1.
DR Pfam: PF02984; cyclin_C; 1.
DR SMART: SM00385; CYCLIN; 2.
DR PROSITE: PS00292; CYCLINS; 1.
KW Cyclin; Cell cycle; Cell division; Mitosis.
SQ SEQUENCE 429 AA; 47900 MW; FE6FF1016FF06253 CRC64;

Query Match 48.9%; Score 43; DB 1; Length 429;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 MKVSRVLAIVLGVIPAH 18
DB 283 MKILRVNFSLGRLPLH 300
II: ||| | | |
|: ||| | | |

RESULT 9
CG1L_MOUSE STANDARD; PRT; 430 AA.
AC P24860;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE G2/mitotic-specific cyclin B1.
GN CCNB1 OR CYCB1 OR CYCB OR CCN-2 OR CCNB1-NS1.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss Webster; TISSUE=Testis;
RX MEDLINE=93080989; PubMed=1280449;
RA Chapman D.L., Wolgemuth D.J.;
RT "Identification of a mouse B-type cyclin which exhibits
RT developmentally regulated expression in the germ line."
RL MOL. Reprod. Dev. 33:259-269(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/He;
RX MEDLINE=92084156; PubMed=1836195;
RA Paterno G.D., Downs K.M.;
RT "Sequence of a cDNA encoding a mouse cyclin B protein."
RL Gene 108:315-316(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92371993; PubMed=1387105;
RA Hanley-Hyde J., Mushinski J.F., Sadofsky M., Huppi K., Krall M.,
RA Kozak A.A., Mock B.;
RT "Expression of murine cyclin B1 mRNAs and genetic mapping of related
RT genomic sequences."
RL Genomics 13:1018-1030(1992).
CC -1- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M

```



```

CC (MITOSIS) TRANSITION.
CC -1- SUBUNIT: INTERACTS WITH THE CDC2 PROTEIN KINASE TO FORM A
CC SERINE/THREONINE KINASE HOLOENZYME COMPLEX ALSO KNOWN AS
CC MATURATION PROMOTING FACTOR (MPF). THE CYCLIN SUBUNIT IMPARTS
CC SUBSTRATE SPECIFICITY TO THE COMPLEX.
CC -1- DEVELOPMENTAL STAGE: ACCUMULATES STEADILY DURING G2 AND IS
CC ABRUPTLY DESTROYED AT MITOSIS.
CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64713; CAA5968.1; -
DR EMBL: X58708; CAA41545.1; -
DR EMBL: S43105; AAB22970.1; -
DR PIR: JH0509; JH0509.
DR PIR: S14652; S14652.
DR HSSP: P30274; IVIN.
DR MCD: MGI:88298; Ccnb1-rs1.
DR InterPro: IPR004366; Cyclin.
DR InterPro: IPR004367; Cyclin_Cterm.
DR Pfam: PF00134; cyclin_1.
DR Pfam: PF02984; cyclin_C; 1.
DR SMART: SM00385; CYCLIN; 2.
DR PROSITE: PS00292; CYCLINS; 1.
DR Cyclin; Cell cycle; Cell division; Mitosis.
KM DOMAIN 51 85
FT CONFLICT 25 25 K -> M (IN REF. 1).
FT CONFLICT 59 59 R -> T (IN REF. 2).
FT CONFLICT 130 130 S -> C (IN REF. 2).
FT CONFLICT 139 139 L -> P (IN REF. 2).
FT CONFLICT 157 157 D -> S (IN REF. 2).
FT CONFLICT 241 241 L -> I (IN REF. 2).
FT CONFLICT 257 257 E -> D (IN REF. 2).
FT CONFLICT 315 315 E -> R (IN REF. 2).
FT CONFLICT 331 331 Y -> C (IN REF. 1).
FT CONFLICT 340 340 QIMA -> RAFS (IN REF. 2).
FT CONFLICT 351 351 K -> E (IN REF. 2).
FT CONFLICT 419 419 L -> H (IN REF. 2).
SQ SEQUENCE 430 AA; 48052 MW; F3BC9C856F6986E CRC64;

Query Match 48.9%; Score 43; DB 1; Length 430;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

```

```

RA Court D.A., Bertrand H.;
RA "Genetic organization and structural features of maranhar, a
RA RT senseless-inducing linear mitochondrial plasmid of Neurospora
RA crassa.",
RT crassa.",
RL Curr. Genet. 22:385-397(1992).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC [RNA](N).
CC -1- SIMILARITY: BELONGS TO THE PHASE AND MITOCHONDRIAL RNA POLYMERASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55361; CAA39045.1; ALU_SEQ.
DR PIR: S26984; S26984.
DR HSSP: P00573; IARO.
DR InterPro: IPR002092; RNA_pol_phage.
DR Pfam: PF00940; RNA_pol_1.
DR PROSITE: PS00489; RNA_POL_PHAGE; 2; 1.
DR PROSITE: PS00900; RNA_POL_PHAGE; 1; 1.
KM Transferase; Transcription; DNA-directed RNA polymerase;
KM Mitochondrion; Plasmid.
FT ACT_SITE 546 546 BY SIMILARITY.
FT ACT_SITE 617 617 BY SIMILARITY.
FT ACT_SITE 798 798 BY SIMILARITY.
SQ SEQUENCE 896 AA; 103972 MW; 112ED0C68225A4D8 CRC64;

Query Match 48.9%; Score 43; DB 1; Length 896;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jaeger K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittowisch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Wolchert G., Aert R., Robben J., Grynopre B.,
 RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Meuller-Auer S.,
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Laureau V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RA "The genome sequence of Schizosaccharomyces pombe.",
 RA Nature 415:871-880(2002).
 RL Mature 415:871-880(2002).
 RN [3]
 RP SEQUENCE OF 14-161: 636-871, 998-1098 AND 1380-1547 FROM N.A.
 RX MEDLINE=96354912; PubMed=8769419;
 RA Saitoh S., Tanahashi K., Nabeshima K., Yamashita Y., Nakaseko Y.,
 RA Hirata A., Yanagida M.,
 RT "Aberrant mitosis in fission yeast mutants defective in fatty acid
 RT synthetase and acetyl CoA carboxylase.",
 RL J. Cell Biol. 134:949-961(1996).
 CC -1- FUNCTION: THIS PROTEIN CARRIES THREE FUNCTIONS: BIOTIN CARBOXYL
 CC CARRIER PROTEIN, BIOTIN CARBOXYLASE, AND CARBOXYLTRANSFERASE.
 CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3)(-) = ADP + phosphate
 CC + malonyl-CoA.
 CC -1- CATALYTIC ACTIVITY: ATP + biotin-carboxyl-carrier protein + CO(2)
 CC = ADP + phosphate + carboxybiotin-carrier protein.
 CC -1- COFACTOR: BIOTIN.
 CC -1- ENZYME REGULATION: BY PHOSPHORYLATION (BY SIMILARITY).
 CC -1- PATHWAY: Long-chain fatty acid biosynthesis; first (rate limiting)
 CC step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: PARTIAL TO CARBAMOYL PHOSPHATE SYNTHETASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: D78169; BAA11238.1; -
 DR EMBL: Z59261; CAB16395.1; -
 DR EMBL: D83413; BAA11914.1; -
 DR EMBL: D83414; BAA11915.1; -
 DR EMBL: D83416; BAA11917.1; -
 DR EMBL: D83415; BAA11916.1; -
 DR HSSP: P24182; 1DV1.
 DR InterPro: IPR001882; Biotin_attach.
 DR InterPro: IPR000089; Biotin_1lpoyl.
 DR InterPro: IPR000901; CPSase.
 DR InterPro: IPR000022; Carboxyl_trans.
 DR Pfam: PF00289; CPSase_L_chain.1.
 DR Pfam: PF00364; biotin_1lpoyl.1.
 DR Pfam: PF01039; Carboxyl_trans.1.
 DR Pfam: PF02785; Biotin_carb.C.1.
 DR Pfam: PF02786; CPSase_L_D2.1.
 DR PROSITE: PS00188; BIOTIN.1.
 DR PROSITE: PS00866; CPSASE.1; 1.
 DR PROSITE: PS00867; CPSASE.2; 1.
 KW Fatty acid biosynthesis; Biotin; Ligase; Multifunctional enzyme;

KW ATP-binding; Phosphorylation.
 FT NP_BIND 266 271
 FT ACT_SITE 393 393
 FT BINDING 745 745
 FT CONFLICT 14 42
 FT CONFLICT 258 258
 FT CONFLICT 339 340
 FT CONFLICT 512 512
 FT CONFLICT 523 523
 FT CONFLICT 636 639
 FT CONFLICT 998 1005
 FT CONFLICT 1017 1017
 FT CONFLICT 1073 1073
 FT CONFLICT 1098 1098
 FT CONFLICT 1105 1105
 FT CONFLICT 1362 1362
 FT CONFLICT 1427 1427
 FT CONFLICT 1444 1444
 FT CONFLICT 1445 1445
 FT CONFLICT 1449 1449
 FT CONFLICT 1451 1451
 FT CONFLICT 1465 1465
 FT CONFLICT 1480 1480
 FT CONFLICT 1485 1485
 FT CONFLICT 1513 1547
 SQ SEQUENCE 2280 AA: 256840 MW: 82629A1A5C8E891 CRC64;
 Query Match 48.9%; Score 43; DB 1; Length 2280;
 Best Local Similarity 53.3%; Pred. No. 87;
 Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 QY 5 RVALVGVAVIPAH 19
 DB 587 KMLAVGALVRAHA 601
 ID SHLE_HUMAN STANDARD; PRT; 365 AA.
 AC P28566; Q9P1Y1;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 5-hydroxytryptamine 1E receptor (5-HT-1E) (serotonin receptor) (5-
 DE HT1E) (S31).
 GN HT1E.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=92302274; PubMed=1608964;
 RX McAllister G., Charlesworth A., Snodin C., Beer M.S., Noble A.J.,
 RA Middlemiss D.N., Iversen L.L., Whiting P.,
 RT "Molecular cloning of a serotonin receptor from human brain (5HT1E):
 RT a fifth 5HT1-like subtype".
 RL Proc. Natl. Acad. Sci. U.S.A. 89:5517-5521(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92128546; PubMed=1733778;
 RA Levy F.O., Gudermann T., Birnbaumer M., Kaumann A.J., Birnbaumer L.,
 RT "Molecular cloning of a human gene (531) encoding a novel serotonin
 RT receptor mediating inhibition of adenylyl cyclase".
 RL FEBS Lett. 296:201-206(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92382553; PubMed=1513320;
 RA Zgombick J.M., Schechter L.E., Macchi M., Hartig P.R.,

RA Branchek T.A., Weinschenk R.L.;
 RT "Human gene S31 encodes the pharmacologically defined serotonin 5-
 RT hydroxytryptamine receptor.";
 RL Mol. Pharmacol. 42:180-185(1992).
 RN
 RP SEQUENCE FROM N.A.
 RA Tromans A.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [15]
 RP SEQUENCE OF 1-363 FROM N.A.
 RA Kitano T., Kobayekawa H., Saitou N.;
 RT Silver Project.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS IS ONE OF THE SEVERAL DIFFERENT RECEPTORS FOR 5-
 CC HYDROXYTRYPTAMINE (SEROTONIN), A BIOGENIC HORMONE THAT FUNCTIONS
 CC AS A NEUROTRANSMITTER, A HORMONE, AND A MITOGEN. THE ACTIVITY OF
 CC THIS RECEPTOR IS MEDIATED BY G PROTEINS THAT INHIBITS ADENYLATE
 CC CYCLASE ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC STRONGEST TO THE OTHER SHT-1 SUBTYPE RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, M91467; AA58353.1; -
 CC EMBL, Z11166; CA477558.1; -
 CC EMBL, M92826; AA58353.1; -
 CC EMBL, AL157777; CAC10582.1; -
 CC EMBL, AB041373; BAA94458.1; -
 CC PIR, S20579; S20579.
 CC PIR, A45260; A45260.
 CC GeneW, HGNC:5291; HTR1E.
 CC MIM: 182132; -
 CC DR InterPro: IPR000276; GPCR_Rhodopsin.
 CC DR Pfam: PF00001; 7tm1.1; 1.
 CC DR PRINTS: PR00237; GPCRHOOPS.
 CC DR PROSITE: PS00237; G-PROTEIN_RECP_FL_1; 1.
 CC DR PROSITE: PS00262; G-PROTEIN_RECP_FL_2; 1.
 CC KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 CC MultiGene family.
 CC FT DOMAIN 1 22
 CC FT TRANSMEM 23 47
 CC FT DOMAIN 48 59
 CC FT TRANSMEM 60 81
 CC FT DOMAIN 82 96
 CC FT TRANSMEM 97 118
 CC FT DOMAIN 119 138
 CC FT TRANSMEM 139 159
 CC FT DOMAIN 160 179
 CC FT TRANSMEM 180 201
 CC FT DOMAIN 202 291
 CC FT TRANSMEM 292 314
 CC FT DOMAIN 315 324
 CC FT TRANSMEM 325 347
 CC FT DOMAIN 348 365
 CC FT CARBOHYD 2
 CC FT CARBOHYD 5
 CC FT CARBOHYD 173
 CC FT DISULFID 95
 CC SO SEQUENCE 365 AA; 41682 MW; 4C31DD783A3F7483 CRC64;

Query Match 47.7%; Score 42; DB 1; Length 365;
 Best Local Similarity 57.1%; Pred. No. 28;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 KVSRLVATLVGAVI 15
 DB 288 KAARILGLIGAFI 301

RESULT 13
 SSUC_BACSU
 ID SSUC_BACSU STANDARD; PRT; 276 AA.
 AC P40401;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative aliphatic sulfonates transport protein ssuc.
 GN SSUC.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BD99 / MS11;
 RX MEDLINE=94281248; PubMed=8011666;
 RA Quirk P.G., Guifanti A.A., Clejan S., Cheng J., Kruhwald T.A.;
 RT Isolation of tn917 insertional mutants of Bacillus subtilis that are
 RT resistant to the protonophore carbonyl cyanide
 RT m-chlorophenylhydrazine.
 RL Biochim. Biophys. Acta 1186:27-34(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98455824; PubMed=9782504;
 RA van der Ploeg J.R., Cummings N.J., Leisinger T., Connerton I.F.;
 RT "Bacillus subtilis genes for the utilization of sulfur from aliphatic
 RT sulfonates."
 RL Microbiology 144:2555-2561(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunze F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borrius R., Boursier L., Briens A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entlin K.D., Errington J., Fabret C., Ferrati E., Fougere D.,
 RA Fritz C., Fujita M., Fujita Y., Funa S., Galliz J., Galleron N.,
 RA Gim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Hatach J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
 RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretto A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis".
 RL Nature 390:249-256(1997).
 CC -1- FUNCTION: PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM FOR
 CC ALIPHATIC SULFONATES. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION
 CC OF THE SUBSTRATE ACROSS THE MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC -1- INDUCTION: REPRESSED BY SULFATE AND CYSTEINE.
 CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
 CC SYSTEM PERMEASE FAMILY. CYSTW SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L16808; AAA64349.1; -
DR EMBL: 293102; CAB07522.1; -
DR EMBL: 293108; CAB2713.1; -
DR Subtilist; BG10657; ssuc.
DR InterPro: IPR000515; BPD.transp.
DR Pfam: PF00528; BPD.transp. 1.
DR PROSITE: PS00402; BPD.TRANSF. INN MEMBR: FALSE_NEG.
KW Transmembrane: Transport; Complete proteome.
FT TRANSMEM 32 52 POTENTIAL.
FT TRANSMEM 54 74 POTENTIAL.
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 242 262 POTENTIAL.
SQ SEQUENCE 276 AA; 30240 MW; FAF16D267419B1D6 CRC64;

Query Match 46.6%; Score 41; DB 1; Length 276;
Best Local Similarity 41.2%; Pred. No. 32;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 2 KVSRLVALVGAIVPAH 18
Db 142 EVSKILLIAGAFEPYV 158
:::1:1111:

RESULT 14
ALP_TRIHA STANDARD; PRT; 409 AA.
AC 003420;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alkaline proteinase precursor (EC 3.4.21.-) (ALP).
GN PRB1.
OS Trichoderma harzianum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Trichoderma.
OX NCBI_TaxID=5544;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=IMI 206040;
RX MEDLINE=9331657; PubMed=8326868;
RA Geremia R.A., Goldman G.H., Jacobs D., Ardiles W., Villa S.B.,
RA van Montagu M., Herrera-Estrella A.;
RT "Molecular characterization of the proteinase-encoding gene, prb1,
RT related to mycoparasitism by Trichoderma harzianum.";
RL Mol. Microbiol. 8:603-613(1993).
CC -1- FUNCTION: SERINE PROTEASE, SECRETED SPECIFICALLY DURING THE
CC MYCOPARASITIC PROCESS, WHICH IS INVOLVED IN THE DEGRADATION OF
CC PHYTOPATHOGEN CELL WALLS, MEMBRANES AND OF THE PROTEINS RELEASED
CC AFTER LYSIS OF THE HOST.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- INDICCTION: BY MYCELIA, FUNGAL CELL WALLS, AND CHITIN, BUT ONLY
CC IN THE ABSENCE OF GLUCOSE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M87518; AAA34211.1; -
DR EMBL: M87516; AAA34209.1; -
DR PIR: S32905; S32905.
DR HSSP: Q99405; IMPF.
DR MEROPS: S08.09A; -
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Hydrolyase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 120 POTENTIAL.
FT CHAIN 121 409 ALKALINE PROTEINASE.
FT ACT_SITE 161 161 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 192 192 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 353 353 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 409 AA; 42271 MW; DDF548BA236C5E63 CRC64;

Query Match 46.6%; Score 41; DB 1; Length 409;
Best Local Similarity 69.2%; Pred. No. 44;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 5 RVLALVGAIVPA 17
Db 5 RRLALVGAIVPA 17
1111111111

RESULT 15
HEXB_ALTSO STANDARD; PRT; 773 AA.
ID HEXB_ALTSO
AC P49007;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Beta-hexosaminidase B precursor (EC 3.2.1.52) (N-acetyl-beta-
DE glucosaminidase) (Beta-glucosidase) (Beta-N-acetylhexosaminidase)
DE (Beta-NHASE).
GN NAG096.
OS Alteromonas sp. (strain O-77).
OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
OC Alteromonas.
OX NCBI_TaxID=29458;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96031592; PubMed=7574618;
RA Tsujibo H., Fujimoto K., Tanno H., Miyamoto K., Kinura Y.,
RA Imada C., Okami Y., Inamori Y.;
RT "Molecular cloning of the gene which encodes beta-N-
RT acetylglucosaminidase from a marine bacterium, Alteromonas sp. strain
RT O-77".
RL Appl. Environ. Microbiol. 61:804-806(1995).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-
CC acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosaminides.
CC -1- SIMILARITY: BELONGS TO FAMILY 20 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: D29665; BAA06136.1; -
DR HSSP: Q54468; IOBA.
DR InterPro: IPR004866; CarDB_Hex.
DR InterPro: IPR001540; GH_20.
DR Pfam: PF00728; Glyco_hydro_20; 1.
DR Pfam: PF02838; Glyco_hydro_20b; 1.
DR Pfam: PF03173; CHB_HEX; 1.

DR PRINTS: PR00738; GLHYDRLASE20.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 773 BETA-HEXOSAMINIDASE B.
FT DISULFID 46 53 BY SIMILARITY.
FT DISULFID 389 397 BY SIMILARITY.
FT DISULFID 496 542 BY SIMILARITY.
FT ACT_SITE 531 531 CATALYTIC ACID (BINDS TO THE GLYCOSIDIC
LINKAGE) (BY SIMILARITY).
SQ SEQUENCE 773 AA: 86760 MW: DBB379885E004642 CRC64;
OY 1 MKVSRVLALVIGAVIPAAH 19
Db 1 MKFNRLMALFGVSSPLA 19

Search completed: April 26, 2003, 13:36:09
Job time: 1.7264 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:31:13 ; Search time 1.29391 Seconds
(Without alignments)
3025.639 Million cell updates/sec

Title: US-10-026-994-3
Perfect score: 88
Sequence: 1 MKVSRVLALVIGAVIPAH 19

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	59.1	339	16	083186
2	47	53.4	202	6	09TUR5
3	46	52.3	129	6	09TUR1
4	46	52.3	474	16	08R800
5	45	52.3	486	16	09AAT9
6	45	51.1	172	13	09DDA1
7	44	50.0	95	16	08YCX9
8	44	50.0	342	5	09NR907
9	44	50.0	436	2	09FA01
10	44	50.0	483	17	098008
11	44	50.0	707	10	09ZRA6
12	44	50.0	1662	2	P71431
13	43	48.9	129	6	09TUR3
14	43	48.9	202	6	09TUR4
15	43	48.9	329	2	054015
16	43	48.9	485	5	018063

17	43	48.9	540	5	018061	018061 caenorhabd
18	43	48.9	816	5	09BK08	09BK08 toxoplasma
19	43	48.9	816	5	09BK07	09BK07 toxoplasma
20	43	48.9	896	8	09XM74	09XM74 neurospora
21	42	47.7	249	13	093232	093232 cynops pyr
22	42	47.7	250	16	09RK82	09RK82 streptomyc
23	42	47.7	363	6	09N2B6	09N2B6 pan troglod
24	42	47.7	363	6	09N2B5	09N2B5 gorilla gor
25	42	47.7	363	6	09N2B4	09N2B4 pongo pygma
26	42	47.7	395	16	P96675	P96675 bacillus su
27	42	47.7	500	9	080264	080264 vibrio chol
28	42	47.7	1411	16	092K12	092K12 rhizobium m
29	41.5	47.2	505	16	08URM9	08URM9 agrobacteri
30	41	46.6	137	17	09YR01	09YR01 aeropyrum p
31	41	46.6	214	16	09K024	09K024 neisseria m
32	41	46.6	214	16	09J5Y5	09J5Y5 neisseria m
33	41	46.6	262	10	094618	094618 oryza sativ
34	41	46.6	291	10	092V08	092V08 arabidopsis
35	41	46.6	399	2	09RMQ7	09RMQ7 mycobacteri
36	41	46.6	484	16	09KYG5	09KYG5 streptomyc
37	41	46.6	492	16	08YDP8	08YDP8 brucella me
38	41	46.6	955	10	004956	004956 cyanidium c
39	41	46.6	2735	3	08WZX6	08WZX6 neurospora
40	40	45.5	48	2	086261	086261 klebsiella
41	40	45.5	122	16	053953	053953 mycobacteri
42	40	45.5	201	4	09H902	09H902 homo sapien
43	40	45.5	258	2	003534	003534 escherichia
44	40	45.5	260	16	09ZBK7	09ZBK7 streptomyc
45	40	45.5	286	16	09YF04	09YF04 streptococ

ALIGNMENTS

RESULT 1
083186 PRELIMINARY: PRT: 339 AA.

AC 083186;
AT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Conserved hypothetical integral membrane protein.
GN TP0151.
OS Treponema pallidum.
OC Bacteria: Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_Taxid-160;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-NICHOLS;
RX MEDLINE-98332770; PubMed-9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uteback T.,
RA McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT *Complete genome sequence of Treponema pallidum, the syphilis
spirochete.*;
RT Science 281:375-388(1998).
DR EMBL; AE001199; AAC65135.1; -
KW TIGR; TP0151; -
SQ Complete proteome.
SEQUENCE 339 AA: 36905 MW; E40CAFA05CF53349 CRC64;

Query Match 59.1%; Score 52; DB 16; Length 339;
Best Local Similarity 66.7%; Pred. No. 2.6;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 SRVLALVIGAVIPAH 18
DB 74 SLVGLTIGAVIPAH 88

RESULT 2

09TUF5 PRELIMINARY: PRT: 202 AA.
 ID 09TUF5: 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 20, Last annotation update)
 DE Putative G protein-coupled receptor (Fragment).
 GN GPR8.
 OS Cercopithecus aethiops (Green monkey) (Griivet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
 OC Cercopithecinae; Cercopithecus.
 OX NCBI_TaxID=9534;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=99337802; PubMed=10407191.
 RA Lee D.R., Nguyen T., Porter C.A., Cheng R., George S.R., O'Dowd B.F.;
 RT "Two related G protein-coupled receptors: the distribution of GPR7 in
 rat brain and the absence of GPR8 in rodents."
 RL Brain Res. Mol. Brain Res. 71:96-103(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF042362; AAF08332.1; -
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR Pfam: PF00001; 7tm_1; 1.
 DR PRINTS: PR00237; GPCR_Rhodopsn.
 DR PROSITE: PS00237; G-PROTEIN_RECEP_FL1; 1.
 DR PROSITE: PS50262; G-PROTEIN_RECEP_FL2; 1.
 KM G-protein coupled receptor; Glycoprotein; Receptor; Transmembrane.
 FT NON_TER 1
 KW NON_TER 1
 SO SEQUENCE 202 AA; 22601 MW; D4E48BC2A674DBE CRC64;

Query Match 53.4%; Score 47; DB 6; Length 202;
 Best Local Similarity 66.7%; Pred. No. 10;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

RESULT 3

09TUF1 PRELIMINARY: PRT: 129 AA.
 ID 09TUF1: 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Putative G protein-coupled receptor (Fragment).
 GN GPR8.
 OS Syllivagus floridanus (cottontail rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Syllivagus.
 OX NCBI_TaxID=9988;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=99337802; PubMed=10407191.
 RA Lee D.R., Nguyen T., Porter C.A., Cheng R., George S.R., O'Dowd B.F.;
 RT "Two related G protein-coupled receptors: the distribution of GPR7 in
 rat brain and the absence of GPR8 in rodents."
 RL Brain Res. Mol. Brain Res. 71:96-103(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL: AF042362; AAF08332.1; -
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR Pfam: PF00001; 7tm_1; 1.
 DR PROSITE: PS50262; G-PROTEIN_RECEP_FL1; 1.
 KM Receptor.
 FT NON_TER 1
 KW NON_TER 1
 SO SEQUENCE 129 AA; 14240 MW; C9308CC1433CA966 CRC64;

Query Match 52.3%; Score 46; DB 6; Length 129;
 Best Local Similarity 60.0%; Pred. No. 9.6;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 KVSRLVAVLGAVIP 16
 DB 68 KASRYTVLVGFVVP 82

RESULT 4

09R8C0 PRELIMINARY: PRT: 474 AA.
 ID 09R8C0: 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE S-layer homology domain.
 GN TPE2087.
 OS Thermoanaerobacter tengcongensis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
 OX NCBI_TaxID=119072;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-MBAT / JCM11007;
 RX MEDLINE=21992816; PubMed=11997336;
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;
 RT "A complete sequence of T. tengcongensis genome."
 RL Genome Res. 12:689-700(2002).
 DR EMBL: AE013156; AAM25260.1; -
 KM Complete proteome.
 SO SEQUENCE 474 AA; 53073 MW; B373A46B9E5B991E CRC64;

Query Match 52.3%; Score 46; DB 16; Length 474;
 Best Local Similarity 71.4%; Pred. No. 33;
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 5

09AAT9 PRELIMINARY: PRT: 486 AA.
 ID 09AAT9: 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein CC0505.
 GN CC0505.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Pollock I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Ueberback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus."
 DR Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL: AE005723; AAK22492.1; -
 KW TIGR; CC0505; -
 SO Hypothetical protein; Complete proteome.

SQ SEQUENCE 486 AA; 52603 MW; 489FA9EADA379BE0 CRC64;

Query Match Best Local Similarity 52.3%; Score 46; DB 16; Length 486;

Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 6 VLALVLCGAVIPAH 19

DB 11 VLALVLCGAVIPAH 24

RESULT 6

O9DDAI PRELIMINARY; PRT; 172 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 21, Last sequence update)

DE Cyclin B1 (Fragment).

OS Xenopus tropicalis (Western clawed frog) (*Xenopus tropicalis*).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OC NCBI_TaxID=8364;

RP SEQUENCE FROM N.A.

RC TISSUE=EGG;

RA Hochegger H., Klotzbecher A., Kirk J., Howell M., Le Guellec K.,

RA Fletcher K., Duncan T., Schall M., Hunt T.,

RT "New B-type cyclin synthesis is required between meiosis I and II

during Xenopus oocyte maturation."

RT development 0:0-0(0).

CC -1 SIMILARITY: BELONGS TO THE CYCLIN FAMILY.

DR EMBL: AJ303451; CAC22294.1; -

DR HSSP: P30274; IYIN.

DR InterPro: IPR004367; Cyclin.

DR InterPro: IPR004367; Cyclin_Cterm.

DR Pfam: PF00134; cyclin; 1.

DR Pfam: PF02984; cyclin; 1.

DR SMART: SM00385; CYCLIN; 2.

DR PROSITE: PS00292; CYCLINS; 1.

DR Cell cycle; Cell division; Cyclin.

FT NON_TER 1 1

FT NON_TER 172 172

SO SEQUENCE 172 AA; 19901 MW; 3C163DDE2F73C00D CRC64;

Query Match Best Local Similarity 51.1%; Score 45; DB 13; Length 172;

Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 MKVSRVLCGAVIPAH 18

DB 87 MKVSRVLCGAVIPAH 104

RESULT 7

O8YGX9 PRELIMINARY; PRT; 95 AA.

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Outer membrane protein E.

GN BME11026.

OS *Brucella melitensis*.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Brucellaceae; Brucella.

OX NCBI_TaxID=29459;

RP SEQUENCE FROM N.A.

RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;

RX MEDLINE=20020109; PubMed=11756688;

RA Delvecchio V.G., Kapatriel V., Redkar R.J., Patra G., Mijar C., Los T.,

RA Ivanova N., Anderson I., Bhattacharyya A., Lykdis A., Reznik G.,

RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,

RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,

RA Haselkorn R., Kyriades N., Overbeek R.,

RT "The genome sequence of the facultative intracellular pathogen

Brucella melitensis."

RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).

DR EMBL: AE009542; AAL52207.1; -

DR Complete proteome.

SO SEQUENCE 95 AA; 10176 MW; F107BCB4D9C31845 CRC64;

Query Match Best Local Similarity 50.0%; Score 44; DB 16; Length 95;

Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 MKVSRVLCGAVIPAH 19

DB 1 MKVSRVLCGAVIPAH 19

RESULT 8

O9N907 PRELIMINARY; PRT; 342 AA.

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Possible deoxyhypusine synthase.

GN CHRI.66.

OS *Trypanosoma brucei*.

OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OC NCBI_TaxID=5691;

RP SEQUENCE FROM N.A.

RC STRAIN=TREU927;

RA Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,

RA Leonard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,

RA Gerrard C., Rajandream M.A., Barrell B.G.;

RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL: AL359782; CAB95376.1; -

DR HSSP: P49366; IDHS.

DR InterPro: IPR002773; DS.

DR Pfam: PF01916; DS; 1.

DR Prodom: PD007730; DS; 1.

SO SEQUENCE 342 AA; 37378 MW; 1A6BF07664635C44 CRC64;

Query Match Best Local Similarity 50.0%; Score 44; DB 5; Length 342;

Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 3 VSRVLCGAVIPAH 18

DB 251 VSRVLCGAVIPAH 266

RESULT 9

O9FA01 PRELIMINARY; PRT; 436 AA.

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Inulin fructotransferase precursor (EC 2.4.1.93).

GN IFT.

OS *Arthrobacter* sp. A-6.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Micrococcales; Micrococcaceae; *Arthrobacter*.

OX NCBI_TaxID=121384;

RP SEQUENCE FROM N.A.

RC STRAIN=A-6;

RX Kim H.-Y., Choi Y.-J.;

RA "Molecular cloning and expression of the thermostable inulin

RT fructotransferase (depolymerizing) gene of *Arthrobacter* sp. A-6 in
 RT Escherichia coli."
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AP124980; AAC09641.1; -
 KW Glycosyltransferase; Signal; Transferase.
 FT SIGNAL
 SQ SEQUENCE 436 AA; 46116 MW; 0A8B519089250A5 CRC64;

Query Match 50.0%; Score 44; DB 2; Length 436;
 Best Local Similarity 75.0%; Pred. No. 63;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 8 ALVLGAVIPAH 19
 DB 20 ALALGAAAPAH 31

RESULT 10
 O980G8 PRELIMINARY; PRT: 483 AA.

AC Q980G8; 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE NADH dehydrogenase subunit N (NUON).
 GN NUON OR S500329.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 CC Sulfolobus
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE-21332296; PubMed-11427726;
 RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aways M.J., Chan-Welher C.C.T., Clausen I.G., Curtis B.A.,
 RA De Moors A., Erasmo G., Fletcher C., Gordon P.M.K.,
 RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault G., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 DR EMBL: AE006667; AA040664.1; -
 DR InterPro: IPR001911; Na/dhco_symp.
 DR InterPro: IPR003918; NADHbc_oxred4.
 DR InterPro: IPR001750; Oxidored_g1.
 DR Pfam: PF00361; Oxidored_g1.1.
 DR PRINTS: PR00173; EDTNSPORT.
 DR PRINTS: PR01437; NUOXDRDTASE4.
 KW Complete proteome.
 SQ SEQUENCE 483 AA; 53086 MW; 79A625B9585DA000 CRC64;

Query Match 50.0%; Score 44; DB 17; Length 483;
 Best Local Similarity 57.1%; Pred. No. 70;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 VSRVLAVALGAVIP 16
 DB 459 VSAIILSIILGIYIP 472

RESULT 11
 O92RA6 PRELIMINARY; PRT: 707 AA.

AC Q92RA6; 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Cobalamln independent methionine synthase.
 GN MS1.
 OS Chlamydomonas moewusii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3054;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Henskens H., den Hartog M.,
 RT "A methionine synthase."
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U77388; AAD00267.1; -
 DR InterPro: IPR002629; Methionine_synt.
 DR Pfam: PF01717; Methionine_synt_1
 SQ SEQUENCE 707 AA; 76264 MW; BDEA93924433C07F CRC64;

Query Match 50.0%; Score 44; DB 10; Length 707;
 Best Local Similarity 61.5%; Pred. No. 1e+02;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 6 VLALVIGAVIPAH 18
 DB 206 LIALLAGAVLPCH 218

RESULT 12
 P71431 PRELIMINARY; PRT: 1662 AA.

AC P71431; 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MoFA protein precursor.
 GN MOFA.
 OS Lepidoptrix discophora.
 OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
 OC Lepidoptrix.
 OX NCBI_TaxID=89;
 RN [1]
 RP SEQUENCE OF 1-1150 FROM N.A.
 RC STRAIN-SS-1;
 RA Corstjens P.L.;
 RL Thesis (1993), Biochemistry, Leiden University.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SS-1;
 RA Corstjens P.L.A.M., de Vrind J.P.M., Goosen T., de Vrind-de Jong E.W.;
 RT "Identification and Molecular Analysis of the Lepidoptrix discophora
 RT SS-1 moFA Gene, a Gene Putatively Encoding a Manganese Oxidizing
 RT Protein with Copper Domains."
 RL Geomicrobiol. J. 14:91-108(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SS-1;
 RA Corstjens P.L.;
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: 225774; CA81037.1; -
 DR InterPro: IPR001865; Ribosomal_S2.
 DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_1.
 KW Signal.
 FT SIGNAL
 FT CHAIN 1 1662 POTENTIAL.
 SQ SEQUENCE 1662 AA; 174293 MW; 639FE238600D9246 CRC64;

Query Match 50.0%; Score 44; DB 2; Length 1662;
 Best Local Similarity 64.3%; Pred. No. 2.2e+02;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 6 VLALVIGAVIPAH 19
 DB 18 LIALLAGAVLPCH 31

RESULT 13
 O9TUF3 PRELIMINARY; PRT: 129 AA.

AC O9TUF3; 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Cobalamln independent methionine synthase.
 GN MS1.
 OS Chlamydomonas moewusii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

```

DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE Putative G protein-coupled receptor (Fragment).
GN GPR8.
OS Cynocephalus variegatus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Dermoptera; Cynocephalidae; Cynocephalus.
OX NCB1_TaxID=9457;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99337802; PubMed=10407191;
RA Lee D.K., Nguyen T., Porter C.A., Cheng R., George S.R., O'Dowd B.F.;
RT "Two related G protein-coupled receptors: the distribution of GPR7 in
RT rat brain and the absence of GPR8 in rodents."
RL Brain Res. Mol. Brain Res. 71:96-103(1999).
DR EMBL: AF042364; AAF08334.1; -.
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECEP_F1_2; 1.
KN Receptor.
FT NON_TER
SQ SEQUENCE 129 AA; 14418 MW; 035B7A355B012A55 CRC64;

Query Match
Best Local Similarity 48.9%; Score 43; DB 6; Length 129;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 KVSRLALVLGAVIP 16
DB 68 KASRIYTLVLGEVWP 82

RESULT 14
O9TUF4 PRELIMINARY; PRT; 202 AA.
AC O9TUF4;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
DT 01-MAY-2000 (TREMblrel. 20, last annotation update)
DE Putative G protein-coupled receptor (Fragment).
GN GPR8.
OS Propithecus verreauxi (White sifaka).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Strepsirrhini; Indridae; Propithecus.
OX NCB1_TaxID=34625;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99337802; PubMed=10407191;
RA Lee D.K., Nguyen T., Porter C.A., Cheng R., George S.R., O'Dowd B.F.;
RT "Two related G protein-coupled receptors: the distribution of GPR7 in
RT rat brain and the absence of GPR8 in rodents."
RL Brain Res. Mol. Brain Res. 71:96-103(1999).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL: AF042363; AAF08333.1; -.
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCR_RHODOPSIN.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Glycoprotein; Receptor; Transmembrane.
FT NON_TER
SQ SEQUENCE 202 AA; 22622 MW; 98D6B5B409B3F18E CRC64;

Query Match
Best Local Similarity 48.9%; Score 43; DB 6; Length 202;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 KVSRLALVLGAVIP 16
DB 68 KASRIYTLVLGEVWP 82

```

```

DB 119 QASRYTTLVLGEVWP 133

RESULT 15
O54015 PRELIMINARY; PRT; 329 AA.
AC O54015;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE Abca protein.
GN ABCA.
OS Paracoccus denitrificans.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Paracoccus.
OX NCB1_TaxID=266;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PD122;
RX MEDLINE=20576169; PubMed=11133961;
RA Harms N., Reijnders W.N.M., Koning S., van Spanning R.J.M.;
RT "Two-Component system that regulates methanol and formaldehyde
RT oxidation in Paracoccus denitrificans."
RL J. Bacteriol. 183:664-670(2001).
DR EMBL: AJ223460; CA11379.1; -.
DR J. Bacteriol. 183:664-670(2001).
SQ SEQUENCE 329 AA; 34788 MW; 266ECC6AE27C358D CRC64;

Query Match
Best Local Similarity 48.9%; Score 43; DB 2; Length 329;
Matches 10; Conservative 4; Mismatches 7; Indels 6; Gaps 1;

OY 1 MKVSRLALVLGA-----VIPAH 18
DB 1 MKISRRITLLAGAAALAGLALPAH 24

```

Search completed: April 26, 2003, 13:37:19
Job time : 2.29391 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 06:44:34 ; Search time 4303.99 Seconds
(without alignments)
17019.508 Million cell updates/sec

Title: US-10-026-994-4

Perfect score: 2517
Sequence: 1 atgaagctctctcagctcctc.....attactaccagctgtgtgtga 2517

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*
1: gb-da:*
2: gb-hcg:*
3: gb-in:*
4: gb-om:*
5: gb-ov:*
6: gb-pat:*
7: gb-ph:*
8: gb-pl:*
9: gb-pr:*
10: gb-ro:*
11: gb-sts:*
12: gb-sy:*
13: gb-un:*
14: gb-vi:*
15: em-da:*
16: em-fun:*
17: em-hum:*
18: em-in:*
19: em-mu:*
20: em-om:*
21: em-or:*
22: em-ov:*
23: em-pat:*
24: em-ph:*
25: em-pl:*
26: em-ro:*
27: em-sts:*
28: em-un:*
29: em-vi:*
30: em-hcg-hum:*
31: em-hcg-inv:*
32: em-hcg-other:*
33: em-hcg-mus:*
34: em-hcg-pln:*
35: em-hcg-rod:*
36: em-hcg-mam:*
37: em-hcg-vit:*
38: em-sy:*
39: em-htgo-hum:*
40: em-htgo-mus:*
41: em-htgo-other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	623.8	24.8	2745	8	AB015511	AB015511 Aspergill
2	481.4	19.1	2579	8	AB1292929	AB1292929 Agarticus
3	436.2	17.3	41906	1	SC5C7	AL031515 Streptomy
4	334.6	13.3	3959	8	AY040839	AY040839 Aspergill
5	243	9.7	3262	1	AF078038S1	AF078038 Caldicell
6	203.6	8.1	12732	1	AE007608	AE007608 Clostridi
7	127.4	5.1	3030	8	VCA429230	AL429230 Volvox ca
8	116.8	4.6	11910	1	AE012276	AE012276 Xanthomon
9	105.4	4.2	14520	1	AE011809	AE011809 Xanthomon
10	98.6	3.9	58930	2	AC098321	AC098321 Rattus no
11	93	3.7	88203	5	AC097628	AC097628 Takifugu
12	92.8	3.7	232869	10	AL603925	AL603925 Mouse DNA
13	92.2	3.7	16150	1	MTV044	AL021999 Mycobacte
14	92.2	3.7	16168	1	AE006985	AE006985 Mycobacte
15	92	3.7	181179	9	AL161725	AL161725 Human DNA
16	91.2	3.6	15311	1	AE007164	AE007164 Mycobacte
17	89.4	3.6	87076	9	AC005918	AC005918 Homo sapi
18	88	3.5	15311	1	AE007164	AE007164 Mycobacte
19	87.8	3.5	126599	2	AP003816	AP003816 Oryza sat
20	87.8	3.5	167254	9	CNS05TDS	AL357093 Human chr
21	87.2	3.5	65140	6	AX211705	AX211705 Sequence
22	87.2	3.5	123580	1	AF263912	AF263912 Streptomy
23	86.2	3.5	125401	6	AX211739	AX211739 Sequence
24	86.2	3.4	200350	2	AC095469	AC095469 Rattus no
25	86	3.4	169585	2	AC078821	AC078821 Homo sapi
26	85.8	3.4	14860	1	AE007093	AE007093 Mycobacte
27	85.8	3.4	63033	1	MTV008	AL021246 Mycobacte
28	85.8	3.4	154292	8	AP003437	AP003437 Oryza sat
29	85.8	3.4	155574	2	AC091090	AC091090 Homo sapi
30	85.6	3.4	47852	1	MTV023	AL022022 Mycobacte
31	85.2	3.4	347750	1	AP002998	AP002998 Mesorhizo
32	85	3.4	15348	1	AE007163	AE007163 Mycobacte
33	84.8	3.4	42655	1	SC7H2	AL109732 Streptomy
34	84.8	3.4	47852	1	MTV023	AL022022 Mycobacte
35	81.2	3.2	120185	8	AC007138	AC007138 Arabidops
36	81.2	3.2	198220	8	ATCHR1V5	AL161493 Arabidops
37	80	3.2	157922	2	AC123764	AC123764 Homo sapi
38	79.8	3.2	20701	1	AE006958	AE006958 Mycobacte
39	79.8	3.2	29550	1	MTV039	AL021942 Mycobacte
40	79.8	3.2	40740	9	AC027349	AC027349 Homo sapi
41	77.6	3.1	167390	9	AC007263	AC007263 Homo sapi
42	77.4	3.1	38425	1	SCF55	AL132991 Streptomy
43	76.8	3.1	1833	6	A37836	A37836 Sequence 6
44	76.8	3.1	1833	6	AR069874	AR069874 Sequence
45	76.8	3.1	1833	6	AR099270	AR099270 Sequence

ALIGNMENTS

RESULT 1
AB015511
LOCUS AB015511 2745 bp mRNA linear PLN 20-JUN-1998
DEFINITION Aspergillus aculeatus mRNA for Avicelase III, complete cds.
ACCESSION AB015511
VERSION AB015511.1 GI:3242654
KEYWORDS Avicelase III.
SOURCE Aspergillus aculeatus
ORGANISM Aspergillus aculeatus CDNA to mRNA.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE 1 (bases 1 to 2745)
AUTHORS Aral,M., Takada,G., Kawaguchi,T. and Sumitani,J.
TITLE Avicelase III from Aspergillus aculeatus
JOURNAL Published Only in Database (1998)

REFERENCE 2 (bases 1 to 2745)
 AUTHORS Arai, M., Takada, G., Kawaguchi, T. and Sumitani, J.
 TITLE Direct Submission
 JOURNAL Submitted (12-JUN-1998) Motoo Arai, Osaka Prefecture University,
 College of Agriculture, Gakuencho 1-1, Sakai, Osaka 599-8531, Japan
 (E-mail: motoo@biochem.osaka-u.ac.jp, Tel: 81-722-54-9465,
 Fax: 81-722-54-9465)

FEATURES
 source Location/Qualifiers
 1. 2745
 /organism="Aspergillus aculeatus"
 /db_xref="taxon:5053"
 1. 2745
 /gene="avIIIIT"
 8. 2578
 /gene="avIIIIT"
 /product="avIIIITase III"
 /codon_start=1
 /protein_id="BA29031.1"
 /db_xref="GI:3242655"

gene
 CDS

BASE COUNT 571 a 803 c 820 g 551 t
 ORIGIN
 Query Match 24.8% Score 623.8; DB 8: Length 2745;
 Best Local Similarity 57.2% Pred. No. 1.7e-85;
 Matches 1354; Conservative 0; Mismatches 957; Indels 56; Gaps 10;

24 CCTTGTCTGGGGCGCCGTCATCCCTGCTCCCTTTTCATGAGAACGTCAGCT 83
 46 CTTGGGCAAGCTTGGACAGAGCTGCGGCTGCGAGGCTACAGTGGAGAACGTCGTCAC 105
 84 GGGGGCGGCGGCGCTTGTGCTCCCGGCTCATCTTCCATCCCAAGAACAGCGCTAC 143
 106 GGGTGGCGAGAGGTTTACCCCGGATGCTTCAATCCCTCCGCAAGGCGCTGGC 165
 144 ATATGACAGAACAGATATTGGCGGCGCTACCGGCTCAACGCGGACGACTCATGACCGC 203
 166 GTACCGGCGGCGAGATATGAGAGGCGGCTACCGTCAACGCTGGAACGCTGGACCGC 225
 204 CGTACGAGATGGGATTTGCTGATATGCGGCTGGCACAACGCGGAGCATGACGCTGTC 263
 226 GTTGTATGAGCTGCGGAGCAATGATATAC---CTGGCATGACTGCGGAGCATGATGCCCTTGC 282
 264 GCTTATCCGAGAGAGATCAAAAGGTATAGCCGAGCGGAGCATGATAGCAACAGCGT 323
 283 CACGATCCGCTGCGACAGGATGCTGTGATCGTGGCGGTCATGATATACCAATGAATAG 342
 344 GGATCCGAGTATGAGAGCATATTCGCTGTCAGACGCGGCGGCAACGCTGCTCTTAC 383
 343 GGACCCCAATGAGGATGATCTGCGTCCGACGAGGACGAGGATGATACCTGACATGAGAC 402
 384 CACTTCCCTTCAAAAGTCGGGGGTAACATGCGACGAGCGGAGCGGAGAGCTCTGGC 443
 403 GAAAGCTGCGCTTCAAGGTTGGGGAATATGCCGCTGCGGGAATGGGAGAGATATGGC 462
 444 TGTGATCGGCGCACTCAACATCATCTATTGTTGCTGCTCAGAAAGCGGCTCG 503
 463 GTTCACCGCAACAAGATATGACTTCTACTTTTGGAGCGGAGATGGCGGCTTGTG 522

504 GAAGTACAGAGAGCGGCGGTGACCTTTTCAAGGCTGCTGCTTACGCGCAACTGGGAC 563
 523 GAAGACACAGATATTAGAGCGAGCTGTGATATGATACCTCTTACGTCGACGAGAAC 582
 564 GTACATCCGAGACCGAGTATGATTCACACGCTACACACGACGACGACGACGATATG 623
 583 GATTTTCCAGGATTTGAG-----CTGACATCAACTTGGAGACCCGTTGGATTTGCTGG 636
 624 GGTTCAGCTTGCAGCTACACGACGACGACGCGGAGGACGACGCTCTGATTTTGTGG 683
 637 GGTGACCTTTGATTCGACCTTGTGAGCTCGGGGCTCTGACGAGCTTGTGCTTGG 696
 684 CAGCGCTGATTAACATCATCTGCTTCACTATGATGACGACGAAATGCGGCTCCACGCGAG 743
 697 AGTCGCGGATTCGTGGCA---AGTCGCTTTCAGATTCGGAGATGCGGCTGCTACTTGGGC 753
 744 TGCCTGACCGGGGACGACGAGGAAATATCTT---CCTCAAGAGCGGAAACTGACGACAC 800
 754 TTGGGTGCTGGAGAGCCTCAGATATGCTTTTCGCCGCAAGGGGGCTTGTCCCTGCA 813
 801 AAGAGAGGCTTGTATCTACCTATTCGATGACGACAGGCGGCTATGATGACGACCTTGG 860
 814 GGAGAGAGCGCTGATCATCTCTACGCGAATGCGCTGCTTACGATGACGACCAAGCG 873
 861 CTGAGTGTGAGGTACGACATTCGACGAGGGGAACTTGGAAAGACATCACCCCTGCTCTGG 920
 874 AACAGTCCCAAGTACACATCATCACACGCGAGTGTGACGATATATGATCCACCTCCT 933
 921 ATCAGATCTATCTTGTGCTTGGCGGCTTGGCGCTTCGATTTGCAAAACCGAGAACCT 980
 934 TCGGATACCTACTACGCTACGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 993
 981 TGTGTTGCTTCTTGAATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1040
 994 GATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1053
 1041 TGGGACACATGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1100
 1054 TGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1113
 1101 CAGCATCTTCACTCCCAAGACGCGTGTATCAAGAACACTTATGATGACGAGCGA 1160
 1114 CAGTATGATATACACACGCGCTTGTGATCCAGAGCA-----C 1152
 1161 GTACCGTCCGATGCTCATCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1220
 1153 GACTTGGACCGCAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1212
 1221 CCCAACCGACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1280
 1213 CCCATTTGACTCCACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1272
 1281 TCTCACCACTGAGACGCGGCTCAATGCTCAATGCTCAATGCTCAATGCTCAATGCTCA 1340
 1273 CCTGACCAATTTGGGACTGAGACCAATGCTCAATGCTCAATGCTCAATGCTCAATGCT 1332
 1341 GGAATTTCCGTCGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1400
 1333 GGAGATGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1392
 1401 CGGAGACGACCAAGCTTTCACCTTTGCCAGAGAAAGCACTTGGAGATCCGCGCAGAC 1460
 1393 TGGGACAGCGGCGGTTCTACACTC-----GATCTGAGACGACGAGCGCAACACCA 1443
 1461 GGTCTGGGACAGCGCCACATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1520
 1444 AGCTTACACACGCGGACCTTACGACGACCAACGCGATCTGCTGCTGCTGCTGCTGCTG 1503
 1521 CAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1580
 1504 TTCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1563

OY	1581	CGGGGCGGGAGCTGTGGAGCATGATCTACGGGGCCGACACAGTCCATCAGAACGGCGGCACGGT	1640
Db	1584	CTTCGGATGACTTTGGTATGACAGACTTAGCGCGCTCAACAGCACCGGACCGGAGCGGT	1623
OY	1641	GGCCTATTTCGGGCGAGGGCGACAGCATCTCTGGTGTCGACACCGGCTGTGCCGGTGCACGC	1700
Db	1624	GSCCTTTTCCGGCCATGGGTGACACGGTGCCTTCTCATGTGTCCABCACTGTGGGGCGCTTGGT	1683
OY	1701	CTTCGCATTTCCAGGGGACGCTTTTGCTCCGTCTCGAGCCTTGCCCCGGGGCGCCGTATCGC	1760
Db	1684	GTCCGAAGTCGCGAGGGGACCCTTGTACGTGGGGTTTTCTCTCCCGCTCGGGGGCGTCATCGC	1743
OY	1761	CTTCGCACAAGAAGACCAACAGCGCTTTTACGCGGGCTCCGGATGACCTTTTACGTCAAG	1820
Db	1744	CTCCCAACAAGTCCGACAACACGCTTCTTATGTGTGATTCGGGGCGGCAATCTACGTGTC	1803
OY	1821	CAMGACACACCGGACACAGCTTACCGCGGGGGCCCAAGGCTGGGCGACGCGAGGAGCATCG	1880
Db	1804	GAMGAAACACAGAGGACCTCTTCTTACCAGAAGCCTGTAGCTTGGGGTCCAGACAGCACCTCAA	1863
OY	1881	GGATATCGCTGCTCACCCGACACCGGACCGGCGACGTTGTATGTCTCGACGACAGTGGCAT	1940
Db	1864	GGCCATCGCGGGCGAATCCCTCCATCGCGGGGAGTGTGGGGGTGCGACGACGAAGAAGGCGT	1923
OY	1941	ATTCCGCTCCACAGACTGGGGCAGACCTTTGGCCAGTCTCCACCGCCCTGACCAACAC	2000
Db	1924	CTGGCATTCACAGGATTAACGGTAGTACGTTCCAGCAGATCGGACAGCGGCTCACGGCGGG	1983
OY	2001	CTACCAAGATTCGCCCCTGGGGGTGGGGCTCAAGGCTCGAAGTGGAAACCTG-----TATGCCCT	2054
Db	1984	CTGGAGACTTTGGGTTTGGGAMAAGGCGTCTCCAAGGGGACGTAACGTGTATCTACGGGTT	2043
OY	2055	CGGCGACCGGCCCCGTGAGGGGCTCGGCTCTACGCCACGTGAGACAGCGCGCTCTCTGAC	2114
Db	2044	CTTTCACCATCGACGGTGGCGGGGGGCTGTTCAGAAGCGAGAGACGAGGACCAACTGGGA	2103
OY	2115	GGACATC---CAGGCTCTCCAGGCGCTTGGGCTCATTCGACAGCACCAAGATCGCGGCAG	2171
Db	2104	GGTAGTCTTCGATGCGGTGCGACGAGATTGGGCTCGGGGTCTCCGCAACGTGTGAACGGCGA	2163
OY	2172	CGGGACACGCCGGGGGGAAGT-CTACGTGTGGGACCAAGCGGCCGGGGCGTCTTTACGCTC	2230
Db	2164	CTTCGACAGACTACGGGCGCGGTCTTTCGTGTGGACAGACGGCGGGGCACTTCTACGGCA	2223
OY	2231	AAGGAACCGTCGGCGGGGCGACGAGCGGGACTTTCCTGTGTCGACCAAGCAGAGACACAGCA	2290
Db	2224	ATCCACAGCGGGGAGCTCGCCGGGCGCCACAGCGCACCGGCGACGCGCGAC- GACCACTCTCA	2282
OY	2291	GTACTCTTTCGCCGACGCTCGAGCACACACGCTGAGGTGAGCGTTGTATCCAGACACCGGG	2350
Db	2283	AGACCTCGACGACGGTATCGACGAGCGTTGAAAACACACACAGCAGCAGCGCTCGACACAGT	2342
OY	2351	CTTTCGACGGTCACTTCGTGAGGACCA 2377	
Db	2343	CTTTCAGACCAACCGTGAAGACCA 2369	
RESULT 2			
LOCUS	ABI292929	2579 bp	mRNA linear PLN 08-AUG-2000
DEFINITION	Agaricus bisporus mRNA for CEL6 protein.		
ACCESSION	AJ292929		
VERSION	AJ292929.1	GI:9755896	
KEYWORDS	cel6 gene.		
SOURCE	Agaricus bisporus.		
ORGANISM	Agaricus bisporus		
REFERENCE	Eukaryotes; Fungi; Basidiomycota; Hymenomyces; Homobasidiomycetes;		
AUTHORS	Agaricales; Agaricaceae; Agaricus.		
TITLE	1 (bases 1 to 2579)		
JOURNAL	Moraes-Almora, P. and Thurston, C.F.		
REFERENCE	Unpublished		
	2 (bases 1 to 2579)		
	Molecular analysis of the cellulolytic genes in Agaricus bisporus		

[illegible]

Oy	427	GCACGAGGCTGGCGTGCATGCCGGCACTCCACATCATCTACTTGGTGGCGC	486
Db	472	ATGGGAAGGGCTCGCGTCGCATCCCAATCTGAACAGATCTTATTTGGGGCTGC	531
Oy	487	TCAGGAACGGCCTCTGGAAGTCTACGGACGGCGGCGTACCTTTTCCMAGTCTCGTCG	546
Db	532	AGTGTAAAGGCTCTCGAAGTCAACAATTTGGGTCGGAGCTGGAGCAAGTAAACAAGC	591
Oy	547	TTCAACGGCACTGGGACGTACATCCAGACCCGAGTGAATTCCACGGCTACACAGCGAC	606
Db	592	TTTCACTACACTGGGCTCTTTCGTTCCCGATCATAGATAGACGGCGCTGAATTCGAT	651
Oy	607	AAGAAGACACTAGTGGGTTTACCTTCGATCGATCAACACAGACAGCAGCAGCGGGAGCCAGC	666
Db	652	AAGAATTGGGATCGATGGGTTTACCATTCGATTAAGCGTCAAGGCTCCTCGTTCGGCTACT	711
Oy	667	TCCTGATCTTTTGTGGACGGCTGATTAACATCACTGCTTCAGTCTATGTAGACAGAAAT	726
Db	712	CCCCGTATATTTTGGCGCGTTTGGCAACAAGGGCTCAAAAGCATTTATATTCAGAAAT	771
Oy	727	GGCGGCTCAACGTGGAGTGTGTACGGGGGACGCCAGGAAATACTTTCATCAACAAGCG	786
Db	772	GGCGGTTCTCTTGGAGTGGCGTCGAGGTCAACACACTAGCTTCTCTCCACATRAAGGC	831
Oy	787	AAATCGAGCCAGACAGAAAGGCCCTGTATCTGACCTATTTCCGATGGCACAAGGCCGTAT	846
Db	832	GTTCCTATCCCTCTGTAGAGGGCACTTTATATCACTTACTCCGATGGTGGGCTCTTAC	891
Oy	847	GATGGCACACTTGGCTTCAGTGTGGAGGTACAGCATTTGCAGGGGGAACTTGGAAAGACATC	906
Db	892	GATGGCAACAAGCGGTGCAGTTTACAAGTACACATTTGACATCTGGCGCATGACGTACATA	951
Oy	907	ACCCCTTCTCTGTGATCAGATCTATCTTGGCTTTGGCGGCTTGGCGCTCGATTTGCCAA	966
Db	952	ACACCGGTACAGCGGAAGACGATCTCTTCTTGGTTTGGAGGCTTAGCTTATAGTAAACAG	1011
Oy	967	AAGCCAGGAACCTTGTGTGCTCTTTGAACTCTTGGTGGCCAGATGTCAGCTGTT	1026
Db	1012	AAAAACGGGACTGTCAAGTTTGGACGCCCTCACTCGTGGTGGCCATGATGATCAATATTC	1071
Oy	1027	CGGTGACACGCACTGTGGGACATCATGAGACCCGATCTGGGCTGGCGAGCTATCCGACT	1086
Db	1072	CGCTGCACAAACGGGGGTGCATTTGGACTCCGCTGTGGAGATGGGCAAGTCTACCCGACT	1131
Oy	1087	GAGACCTATTACTACAGCATCTCACTCCCAAGCACGGGTGATCAGAACAACCTTATTC	1146
Db	1132	CTCAACAATAATCTACAGTTTAAACGCTGCCCTGGCGCCCTGGATAGGACCCCAACATCGTT	1191
Oy	1147	GATGTGACGAGCGAGTCAACCGTCCGATGGTTCATCAACGGCGCTGGGTGATGTAG	1206
Db	1192	GATGTCAACCCCG-----GAAATCTTCAAAATGGGTGATGATGGAA	1233
Oy	1207	TCCTCGAGATTGACCCMACCGACAGCAACCACTGGCTCTACGGACCGGATGACAATC	1266
Db	1234	TCACATATTCATGTATCTCTTTCGACTCGAATCATTTGGCTCTACGGCAGTGGAAACTATTC	1293
Oy	1267	TTTGGCGGCCAGATCTTCACCACTGGGACAGCGGCCCAATGTCTCAATCCATCACTG	1326
Db	1294	TATGAAGTGTGTGCTCTTCAATATGGAGAGTGCACACAGTCAACATCAAGTCGCTC	1353
Oy	1337	GCACAGCGCATCGAAGAAATTCCTCGTCCAGAGACTTGGCTCTGCACCCGCGGAGAAGCGAG	1366
Db	1354	GCAGATGTGTGCGAGGAACACTCCGCTCCAAAGCCTATTTTACCTCTCTTCCGAAACGGCT	1413
Oy	1387	CTAATGGCCGAGTCCGAGACGCAACGGCTTCACTTTTGGCAGAGAAACGACTGGGG	1446
Db	1414	CTTGTTTCTGTATTTGGTGTGACTTTGGCGGTT-----TGTCTCAACAACTACTTACCCGTG	1468
Oy	1447	ACATCGCCGAGAGCGTCTGGGGAACGCCACATATGGGCACTCGACAGAGGTGTGACTAC	1506
Db	1469	CACCCAGCGCCCA-----GTTTACCAATTCCCAAGTGTGTGACCTCGGCTGACATGATTTTC	1524
Oy	1507	GCGCGGAACCTCGGTCAAGAGCGTCTGTCGCGGTGGCAACACCGCGCGGACGC-----AA	1560

Db	1575	GCTGGCATTAACCCAGTAAGTATGTCTGTGGCATTTGTTACTGGTGGATAGTACTAGTGGCAAA	1584
QY	1551	CAGGTGGCCATCTCGTCCGACGGCGCGCGACGTCGTAGCATTCAGTACGCGCGACACG	1620
Db	1585	CAGGTGGCCATCTCTTCCGACTATGATGCGCTTACTTGGATATCAGCATTTTGGAGCTCTGAT	1644
QY	1621	TCCATGACGGGGGACGACGTGCTTATTCGGCGCGCGACGACATCTCTGTCGGACG	1680
Db	1645	AATGTTCAAGGGGGGAAGATTGTCATCTCCGTGTGATGTCGCGATCATCTCTGGCGTACA	1704
QY	1681	GCCTGTCCGGCGTGCACGCGCTCGCATGTTCCAGGCGACGCTT	1721
Db	1705	AATGTAACGGTGTGATGTGTGCGGAACACGACGCGACCTT	1745
RESULT 3			
LOCUS	SC5C7	41906 bp	DNA linear BCT 12-MAY-2002
SC5C7/c			
DEFINITION	Streptomyces coelicolor cosmid 5C7.		
ACCESSION	AL031515	AL645882	
VERSION	AL031515.2	GI:20520758	
KEYWORDS	ATP/GTP-binding protein, lyase; narg; nargH; nargI; nargJ; nitrate reductase; oxidoreductase; RNA polymerase sigma factor; secreted cellulase; uvra-1-like protein.		
SOURCE	Streptomyces coelicolor A3(2).		
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
REFERENCE	1 (bases 1 to 41906)		
AUTHORS	Redemach, M., Kleser, H. M., Denapate, D., Eichner, A., Cullum, J., Kinsahl, H., and Hopwood, D. A.		
JOURNAL	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome		
MEDLINE	Mol. Microbiol. 21 (1), 77-96 (1996)		
PUBMED	97000351		
REFERENCE	2 (bases 1 to 41906)		
AUTHORS	Seeger, K. J. and Harris, D.		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 41906)		
AUTHORS	Parkhill, J., Barrell, B. G. and Rajandream, M. A.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-SEP-1998) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK		
COMMENT	On May 9, 2002 this sequence version replaced gi:355990.		

Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.

(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC757.01c (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(42):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (atc)) which is preceded by an upstream

ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 5c7 lies between 1b6 and 4b5 on the AseI-A genomic restriction fragment.

FEATURES

```
source
1. 41906
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 5c7"
2. 110
/note="nominal overlap with cosmid 5c7 from 1 to 109"
complement(143..147)
/note="possible RBS upstream of SC5c7.01c"
337..340
/note="possible RBS upstream of SC5c7.02"
346..2736
/gene="SC5c7.02"
/note="SC06517"
346..2736
/gene="SC5c7.02"
/note="SC5c7.02"
uvrA-like protein, len: 796 aa; similar
to Streptomyces peucetius TR:054827 (EMBL:L76359)
daunorubicin resistance protein drnc (764 aa), fasta
scores; opt: 2051 z-score: 2702.2 E(): 0, 51.1% identity
in 775 aa overlap, and to many uvrA proteins e.g.
uvrA_NERGO excinuclease ABC subunit A (950 aa), fasta
scores; opt: 412 z-score: 1697.0 E(): 0, 34.1% identity in
953 aa overlap. Contains 2x PS00017 Arp /GTP-binding=
match to entry PF00005 ABC_tran, ABC transporters, scores
42.40, E-value 1.7e-10 and 71.20, E-value 2.2e-17"
/codon_start=1
/transl_table=11
/product="uvrA-like protein"
/protein_id="CAA20614.1"
/db_xref="GI:3559992"
/db_xref="SPTRMBL:086699"
/translation="WSSAKRPGTGPQSHVADSHDLINHGARENILKDVSVDPKRR
LVEFTVSGSGKSSLVNTIAESQRLINETYSAFVGQEMPTLARPEVDVLDGTLAI
IVDDQRMGADRSYTGATIDYVAMLRILFSRLGEPRIQSPASVFNPAVSAGATV
ERGNKKAARATEFERGCMTCCEGRTSIDIDQLYDSDSLAGAFITPMKSDSO
MTVOVYAGSGFVDDPKPTREYTEKEKRLDPTIGEPKYVNVNLTIELIKIQSFL
SKDEAMOPHRAFERAVTTTTCPECGTSLSEGRSSKTKISIDACAMEIRDLA
EWRDLTEPSVAPLTLALRDLTDSFVEIGLGLSLDRPAGTSLSGEADQVKMIRLGS
SLDTTYVDEPTVGLPHDIDRMNDLRLDKGNTVLVEHKPEALAIADHVNLG
PGAGTAGCTVCEFTVEELRAADVTGRLHLDRAVLKSVKRPAGALEIRARFTNLO
GVADVPLGVLCVTVGAGSGKSLIHGSVPAGADVVDOSPIKSGRSRNPATYTL
LDPIRKAFKAGVKKALFASNSGCAFTGACGIVTIDLMACVAPRCDCCKRR
OPALYERFGRGRDSEVLAISYDQAEFEFGAERTPAHKTIORLSVGLGILTLGO
PLTUSGGERORLKLATHMGEKGVYVDEPTTGLHADVDEOLLGLDLRLVDAGKSVI
VIEHQAVMAHADMITIDLPGAGHDGGRVFEETPADLVADRSTLTGHLAAVYA"
472..1701
/gene="SC5c7.02"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporters, score 42.40, E-value 1.7e-10"
493..516
/gene="SC5c7.02"
/note="PS00017 Arp/GTP-binding site motif A (P-loop)"
1894..2628
/gene="SC5c7.02"
/note="Pfam match to entry PF00005 ABC_tran, ABC
transporters, score 71.20, E-value 2.2e-17"
1915..1938
/gene="SC5c7.02"
/note="PS00017 Arp/GTP-binding site motif A (P-loop)"
2804..3286
/gene="SC5c7.03"
/note="SC06518"
2804..3286
```

```
misc_feature
2873..3133
/gene="SC5c7.03"
/note="SC5c7.03, probable transcriptional regulator, len:
160 aa; similar in N-terminus to many e.g. RMA-ENCL
transcriptional activator Rma (113 aa), fasta scores;
opt: 175 z-score: 255.3 E(): 5.7e-07, 32.3% identity in 96
aa overlap. C-terminal 50 aa are unique. Contains Pfam
match to entry PF00165 HTH_2, Bacterial regulatory
heli-turn-helix proteins, arac family, score 71.40,
E-value 1.9e-17"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAA20615.1"
/db_xref="GI:3559993"
/db_xref="SPTRMBL:086700"
/translation="WTTEEDLARLRADLDREAYELDPALARVALMSAGHSRS
FRAAYGETYSYLMTRTERIAMALTRDMSVTEVCVAGCTSSFSRTELVGES
PSAYRERHDAGCALPACVAVYTRPVNGERGRTAGRSSESVRNEAAPAPPPVA"
2873..3133
/gene="SC5c7.03"
/note="Pfam match to entry PF00165 HTH_2, Bacterial
regulatory helix-turn-helix proteins, arac family, score
71.40, E-value 1.9e-17"
3292..3708
/gene="SC5c7.04"
/note="SC06519"
3292..3708
/gene="SC5c7.04"
/note="SC5c7.04, possible lyase, similar to e.g.
LGUL_HABIN lactoyglutathione lyase (PC 4.4.1.5) (135 aa),
fasta scores; opt: 124 z-score: 215.2 E(): 9.7e-05, 29.0%
identity in 131 aa overlap"
/codon_start=1
/transl_table=11
/product="putative lyase"
/protein_id="CAA20616.1"
/db_xref="GI:3559994"
/db_xref="SPTRMBL:086701"
/translation="NDISLSCFIAVDHDLARALARYRYVLGLEAVNDYEGSMKAVTL
GSPADPDIVLEPLADPNASADKQAMADLALGHLGTLTADVDATFEHVRRA
GAEVQEPVDQPYGVRCAPFRDPAGNVRVLRPRKG"
3979..3982
/note="possible RBS upstream of SC5c7.05"
3995..4795
/gene="SC5c7.05"
/note="SC06520"
3995..4795
/gene="SC5c7.05"
/note="SC5c7.05, probable RNA polymerase sigma factor,
len: 266 aa; similar to 5. coelicolor sigma factors of the
SigB subfamily e.g. TR:052313 (EMBL:AF036131) putative RNA
polymerase secondary sigma factor SigG (263 aa), fasta
scores; opt: 1258 z-score: 1540.5 E(): 0, 72.9% identity
in 251 aa overlap, and to RPOF_STRAU RNA polymerase
sigma-F factor (297 aa), fasta scores; opt: 713 z-score:
858.3 E(): 0, 42.6% identity in 312 aa overlap. Possible
alternative start 216 bp upstream, not supported by
frameplot. Contains helix-turn-helix motif at aa 228-249
(sigma70, Sigma-70 factors, score 210.40, E-value 2.8e-59"
/codon_start=1
/transl_table=11
/product="putative RNA polymerase sigma factor"
/protein_id="CAA20617.1"
/db_xref="GI:3559995"
/db_xref="SPTRMBL:086702"
/translation="MPHASKVHPHDDAPPTADARFRLATLPDGEHDELDRIVEAM
LPMARLAGRSGESFEDLRQVALGLVAVDPRYVNAFSPVAVPTTGEIKR
HFRDMMTLVPRRYQDELRNKRYRAGDLSQTIIPRRPTVEIAERALSSEDVAVGL
EALDESFTALSDAEIPGSESDYSLGDALGAPALDVTVDVDEAVRDLAALPERRAI
LYMRFNDMTOSRIAEQISOMHVSRLITRCRDRVRLADPQPA"
```



```

Db 2258 GCAGCACTCACTGCTGCTGCTGATGAGCTGATCTGCGCCAGCAGCATTCCTGGGGC 2317
QY 1048 ACATGAGCCCGATCTGGGCGGAGGCTATCCGACTGAGACTTATCTACTACAGCATC 1107
Db 2318 ACCTGTGACACCATCTGGGCTTGGAGGGCTATCCAGCATCTACTACTACTCTTAT 2377
QY 1108 TCACCTCCCAAGACACCGGTGATCAGAACACTTATCGATGTGAGAGCGACTACCG 1167
Db 2378 GACATCTCCAAAGCCCATGGCTACAGA-----CGACACATCT 2416
QY 1168 TCCGATGCTCTCATCAGAGCCCTCGGCTGATGATGATGCTCTCGAGATTGAGCCCAAC 1227
Db 2417 ACAGAGCAATTTCCCGCTCCGCTCGCTGATGGTGCAGACCCCTCGCATTTGATCCCTTC 2476
QY 1228 GAGAGCAACCACTGGCTCTACGGCAGGAGTGAATCTTTGGCGGCGCAGATCTCAC 1287
Db 2477 GACTCTGACCACTGGCTTACGGCAGCTGGGAGACCATCTACAGCGGCGCATGACCTTCAA 2356
QY 1288 AACTGGGACACGGGCGCACAATGTGTCAATCAATCACTGGCAGAGCGCATCGAGAAATTC 1347
Db 2537 AATTGGGACAGCGAACACAGTAATATCGAATCTTCCCGTAGGCATCGAAGAAATG 2596
QY 1348 TCCGCTCAGACACTGGCTCTGACACCGCGGAGAGCAGCTATTGGCCGAGTCGAGAC 1407
Db 2597 GCTGTCTGGGTCTCATCACTCCCGCAGCGGCGCTGCTCTCTCTGCGGTGCGCGAC 2656
QY 1408 GACACGCGCTTCACTTGGCAGAGAACAGACCTCGGAGATCGCCGCGCAGAGCGTCTGG 1467
Db 2657 GATGGCGGCTT-----CTACACACTTCCCTGACACCGCCCATCATCAATACTAC 2707
QY 1468 GCAACGCGCCACATGGGCGACCTCGAGAGCGGTGAGCTACCGCGGGAATCGGTCAAGAGC 1527
Db 2708 CACACCGCGGCTTACAGTCAACAAAGCGCATGAGTACCGCGGCAACAGCCGCTAAC 2767
QY 1528 GTGTCCGCTCGGCAACACCGCGGCGACCGCAACAGTGGCCATCTGCTCGGAGGCGC 1587
Db 2768 ATGTCCGCTCGTGGATCGACGATTCGACCTCACTCTGCTTATCTTATCTTCAAGTTTGGC 2827
QY 1588 GCGACGTGAGCATGAGTACGCGCGGCGAGCTGATGAACGGCGGCGAGCGTGGCTAT 1647
Db 2828 GAATCTTGGTACCGCGCATGCGCTTCTCTCTCTGAGACCGGCGCAAGTCCGCTC 2887
QY 1648 TCGCGGCGGCGACAGCATCTCTGCTGCTGAGCGGCTGCTCGGCGTGAACGCTCGAG 1707
Db 2888 TCAGCGGCGATCAACACATCTCTCTCATGAACAGC---GACGGTGGCTATCGTTCCGA 2944
QY 1708 TTCCAGGCGGAGCTTGGCTGCTGCTGAGGCTGCGCGGCGCGGCGGCTGATCGCTCGAC 1767
Db 2945 AACTCGGCTACTCTTACGCGAGTCTCTCTGCTGCGCTCTGAGCTGTCATGCGCTCGAC 3004
QY 1768 AAGAGACCAACAGAGCTCTTCTACGCGCGCTCCGATCGACTTTTACGTACAGCAAGAC 1827
Db 3005 AAGGCTAACACACACTACTTCTAGCGAGGCTCCGCGCTTCTTCTTTCATCCGAC 3064
QY 1828 ACCGGCAGCAGCTTACGCGCGGCGGCGCAAGCTGGGCGAGCGGAGGAGATCCGGATATC 1887
Db 3065 ACCGCGCGGCGACTTCACTGTCACACACCGCTCGGCGAGTACACACCGGCAAGCGCATC 3124
QY 1888 GCTGCTACCGGCGACCGCGGCGAGCTTGTATGCTCTGAGCGGAGCTCGGCGCATATTCGC 1947
Db 3125 CCGCGGCGAACCTCTCTCGGAGGCGATGTCTGGTCAGTACAGACAGGCTCTTCTCCAC 3184
QY 1948 TCCACAGACTCGGCGAGCAGACTTTGGCA 1976
Db 3185 TCACCACTAGGCGCAATCTTTCACCCA 3213

```

```

RESULT 5
AF078038S1 3262 bp DNA linear BCT 11-FEB-2001
LOCUS Caldicellulosiruptor sp. Tok7B.1 glycosyl hydrolase 5 gene, partial
DEFINITION cds
ACCESSION AF078038

```

```

VERSION AF078038.1 GI:12743874
KEYWORDS 1 of 5
SEGMENT Caldicellulosiruptor sp. Tok7B.1.
SOURCE Caldicellulosiruptor sp. Tok7B.1
ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales;
          Syntrophomonadaceae; Caldicellulosiruptor.
REFERENCE 1 (bases 1 to 3262)
AUTHORS Gibbs, M.D., Reeves, R.A., Farrington, G.K., Anderson, P.,
          Williams, D.P. and Bergquist, P.L.
TITLE Multidomain and multifunctional glycosyl hydrolases from the
JOURNAL extreme thermophile Caldicellulosiruptor isolate Tok7B.1
MEDLINE Curr. Microbiol. 40 (5), 333-340 (2000)
PUBMED 20171169
REFERENCE 10706665
AUTHORS Gibbs, M.D., Reeves, R.A., Farrington, G.K., Anderson, P.,
          Williams, D.P. and Bergquist, P.L.
TITLE Direct Submision
JOURNAL Submitted (27-OCT-1999) Biological Sciences, Macquarie University,
          Sydney, NSW 2109, Australia
FEATURES
     source
         1..3262
         /organism="Caldicellulosiruptor sp. Tok7B.1"
         /isolate="Tok7B.1"
         /db_xref="taxon:80339"
         /note="ORF5; multidomain"
     CDS
         274..>3262
         /note="ORF5; multidomain"
         /codon_start=1
         /transl_table=11
         /product="glycosyl hydrolase 5"
         /protein_id="AAK06388.1"
         /db_xref="GI:12743879"
         /translation="MKRLSLFSLILLNLFLEPLQIGAKADTGIANOPYWKNKI
EGGGRITIVNPKRNLYVRIDIGGAFIRSTDGNFTQILDWVSFDNMLGVES
IATDPVDPNNXVILIACGTYTNSMTWMLNGAILSTBEGDFETLPFLFXGMPKXNL
GERLADIPNNRNRLYLGTREGXGLMKSDEBYGVSWKVFSEFPNGYVIEDPNCPNYLN
HITGVVWVVEDPTSGRGEKXLIYGVVADKTSIYITADGOTQALPGQPTGLLPQ
RAKLSGDMLYITNTGPNYNDGVEVRYNTKGEKNIIPMAODPTFEYGGIAG
DNQNPVVVAALSSWMPDTYIMRSIDGETKCIEMNGYRNRLIHYVMDI SAAPML
NGXNBPPTPEVSPKIGMAYGTLEIDPESDKMLGTGAXLIGCDDLTWMDGQNTI
KVAKIGIEBTSVALISVPVGPPLSLALDILGFRHEDLEKAPNMTYVOPNMGTTTDI
DPAELNPNMVRGVNDKQNPNTNRIGPSYGGKSMFQGNTEPOGTSGLVVAAD
GSAVWAPKARVAVCYSTDGNKMWBECAVPSBAIYSDRVNPNKEFAFNKGFYISAD
KKKTFIESPAAGLPIISGNFKYTGIEGDIILNNGNMHSTGGVSEFVISEGVEDAS
IGEKPAEGEYPAITYTAKINGVRGIRFSDDCDKTWIRINDKIOFGCANMDIGDP
RYVGRFVATNGLGKMGELIANSISPSATPTSTSEPTPTPTATITSTPASTPT
PTPTPTPTATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
SAGSIRPWKRVNGSSVDSLRSVKIRVWYTVDGKPOSAVGDMAOIGASNVTFNPK
LSSGVSGADYILEVGFSSGAGOLOPKDIDGDIQVRFNKNDMSNYNQADDMWSLOSMTD
YSENAKVTLYVGVGLVWGOEPCGA"
         274..2580
         /note="Region: endoglucanase domain"
         /evidence=experimental
         /misc_feature
             2791..>3262
             /note="type IIIB cellulose binding site"
             2581..2790
             /note="pr-linker"
     BASE COUNT 1056 a 529 c 794 g 871 t 12 others
ORIGIN
Query Match 9.7%; Score 243; DB 1; Length 3262;
Best local Similarity 50.5%; Pred. No. 1.3e-27;
Matches 685; Conservative 8; Mismatches 643; Indels 21; Gaps 4;

```

```

QY 67 TGAAGACGCTCAAGCTGGCGGCGCGGCTGCTCCCGGCAATCATCTTCATCC 126
Db 388 TGAAGATCTTAATAATTGAAGCTGGAGAGTTTATCTAGAAATAGTAAACCCG 447
QY 127 AAGACAAAGCGCTAGCATATGACGACAGATATGGCGGCGGTGAAACGCCCTCA---AC 183
Db 448 AAGAGAGAAATCTTGTACGTTAGGACAGATATGAGAGGAGCACTTACGTAAGAC 507

```


CDS complement(2634..2978)
 /gene="CAC0920"
 /codon_start=1
 /transl_table=11
 /product="Protein related to MifH/DOPD protein family,
 function in bacteria is unknown"
 /protein_id="AAK78896.1"
 /db_xref="GI:15023821"
 /translation="MPFINSTVTVKLDRKESTLVKVELGKTIELDPKSGEDMLWGEK
 DNPVLFGEKKDKAFIEVKIFGSADKASKNKLVAISLSMENELSTIPKDCIYITVD
 EYNNAMNSLF"
 3104..4252
 /gene="CAC0921"
 /gene="CAC0921"
 /codon_start=1
 /transl_table=11
 /product="Metal-dependent amidohydrolase"
 /protein_id="AAK78897.1"
 /db_xref="GI:15023822"
 /translation="MTTIEKOEFLKLKIGYRRELKHPPELSMKEXETTKRIKRLME
 NIKVLPETIEGVVAETIEGHEGKTVARADIDALPIEEKTELPESSVYKGMHACG
 HDHPHAIIGTALISKRAETGPRITTFORGEERKGSQTIIEGVARVDVDCIFGM
 HKKPDLPVGTIGIRSELMAVDREIKVIGSGHAGIPNKCIDPIVASQIVASMQT
 VVSRISPIESVVISITRFNSGTVMNVISDSAGMEGTVFNNMIREEIRNLMERTAK
 GIARALGKIEFKMFSSVPLENDVRFKDCIEADENGKRYVEAKRNIGDGEFSFYQ
 TVPSPFVWVGDTGEEMHKANYNLKEDALMVANYFSDALKVLSKK"
 4594..4863
 /gene="CAC0922"
 /gene="CAC0922"
 /codon_start=1
 /transl_table=11
 /product="Uncharacterized small conserved protein, homolog
 of yjfa/yuke B. subtilis"
 /protein_id="AAK78898.1"
 /db_xref="GI:15023823"
 /translation="MFDKALECEDIKSIKSTQIDISWYGESCTAFKTYLLIDR
 KINTYTEILOEYSKALYDISFTYQNTDGIAMKAMKAVNEASEQK"
 5056..6348
 /gene="CAC0923"
 /gene="CAC0923"
 /codon_start=1
 /transl_table=11
 /product="Predicted membrane protein; CF_1 family"
 /protein_id="AAK78899.1"
 /db_xref="GI:15023824"
 /translation="WQSLKVDNRNGLKLINEAKQGLSAVDKFSKLNKRSISIEILKN
 LKRELKAVNQLGEIRKINSNIDYVKRFTDVKCASLKTSGTYRKVGILNFE
 EVERPVIGIGVPIIDAVESVDNRNFKVTKTTITLETGALGVCIPTLPKAGILNFE
 MDALSAAGVMSLDINIGTVAASAIYORVERKSESDSGDLSIFEGGAGWYSAKILG
 VNPSPYWCSSGKTYLDRTSATVTVNVGSHVSLGKDVKRLPAIKAEASANKTEA
 ANNAKTIHDDAMATKRNKNGINNAKKEVYDRLKLKNYKGTNFEPRYKQFTRKDKNL
 KDKGNKLPQODDLIIQDDQLETFEKNLTNANKNKSIRDLNKNYRDLKASLVNIPA
 GGTNLNTEDDQSSFFHNFPGIKPIFEIKVVLGER"
 6345..6923
 /gene="CAC0924"
 /gene="CAC0924"
 /codon_start=1
 /transl_table=11
 /product="Predicted membrane protein"
 /protein_id="AAK78900.1"
 /db_xref="GI:15023825"
 /translation="MIRQKRAIFPLTISILIGDIEKITSFGIEPFDICNLFPI
 GLGIGFIRIRWYKQIKKYTELLKEDRLDKTEYITIDERVGVYVNTAFYIDGKC
 KLRFPQYCKDNDNEYLEPIYKNGEKFDPYIADKKNDYKKSISTNEITTECEEK
 AEKKRRRLNLYVNSRMLAKKYEENKMK"
 7003..8043
 /gene="CAC0925"
 /gene="CAC0925"
 /codon_start=1
 /transl_table=11
 /product="Cystathionine gamma-synthase"

/codon_start=1
 /transl_table=11
 /product="TPR-repeat-containing protein"
 /protein_id="AAK78901.1"
 /db_xref="GI:15023826"
 /translation="MENENVLRAOELNRLGFIERNAGDEALKYFNNAIKEDENEKE
 AYLNKAALCIAMNKVDVASCYDKLTIKYPNGKAYFGGNGWPFYKDKIKAEIEN
 RSIYLGMESEGVYVNLACMESIGELDDAIKMLNRLIITNDONITLUNKAAVLEIK
 KRVALLCEYNRVLKIEVDNEEYVHKAMVIGMKREKAFETISRGEALGOMTLCY
 DKVAUFKEKREKALECTERKALLDEGVLLILKKGDLITLTKIDKASTYDNLIE
 FHENNMEGVFAKANLCMLGHEHDAEEFFKEITERLSGEESVLIINSYIRALNLKMA
 KTRKQRLIRKO"
 7998..8357
 /gene="CAC0926"
 /gene="CAC0926"
 /codon_start=1
 /transl_table=11
 /product="TPR-repeat-containing protein"
 /protein_id="AAK78902.1"
 /db_xref="GI:15023827"
 /translation="MCKNEAKAEAKYKKNYPLIKYPVVKLNFPKANCLRIOD
 YKAEELIYITDLDNDALLEVYLMKARNRISLNTKAEAKEDLNTYIKLNPAYRIEL
 DQLEKEFLSGSESCYR"
 8379..8675
 /gene="CAC0927"
 /gene="CAC0927"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical protein, CF-2 family"
 /protein_id="AAK78903.1"
 /db_xref="GI:15023828"
 /translation="MNSDEIRIDEKACDVYDKLAVGKYEYTNLIHLSNRHIEDG
 IKGSAVDSITTEYKNIYKWKVKEMQMDGVAVANNIYSSFDAEDMKITKTYR"
 8713..9012
 /gene="CAC0928"
 /gene="CAC0928"
 /codon_start=1
 /transl_table=11
 /product="Uncharacterized small conserved protein, homolog
 of yjfa/yuke B. subtilis"
 /protein_id="AAK78904.1"
 /db_xref="GI:15023829"
 /translation="WVDRCDIFIDASIFETIKYVSDSKVRHLIDLCLENLRKMD
 TWEGDAKRFDSFTFGFYSAMKKKDDMLEELIKELKLVKTSFESLDTEKKNLDKR"
 9712..10335
 /gene="CAC0929"
 /gene="CAC0929"
 /codon_start=1
 /transl_table=11
 /product="SAM-dependent methyltransferase"
 /protein_id="AAK78905.1"
 /db_xref="GI:15023830"
 /translation="MGSMDYFDNIAASNVIRSEYFEERLKYKVLKYSIKDKITADF
 GCGTGFSIALASDANWVGLNSNMVLELKSKSHDKLNNIYILKSLTSTVTFDE
 SIDAVEINMALHIHIDAEKATKEMRYLKKDCTVYISDSSENGCNARREMDENLGF
 SNEQIDMLKLVNGFKNIDIENTDLCKGTSNGEHTERTSIFAKAK"
 10353..11498
 /gene="CAC0930"
 /note="metB"
 /codon_start=1
 /transl_table=11
 /product="Cystathionine gamma-synthase"

Query Match 8.1%; Score 203.6; DB 1; Length 12732;
 Best Local Similarity 50.0%; Pred. No. 1e-21;
 Matches 667; Conservative 0; Mismatches 634; Indels 33; Gaps 5;

33913
/strain="ATCC 33913"
/db_xref="ATCC:33913"
/db_xref="taxon:190485"
/note="pathovar: campestris"
241..1212
/gene="gppr"
/note="XCC1748"
241..1212
/gene="gppr"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="galactose-binding protein regulator"
/protein_id="AA041039.1"
/db_xref="GI:21112841"
/translation="MANSQTPMNAARLKTROLLLLHLEORSVRAEAASMTOPA
ASRLAEMEDMLGVLFERHARGVEPTWGHVLRARAMEISGRAGEIARLRGR
MGOASIGTVNPGTNLPOAIAEVRKEPDIIVRVMDYSRPLAKLGGOLDIVGR
ITGPEGVGELEPELADEPHSYARAGHPLASARLOHADVLRHGWILPPADSVLAR
LOSMMEHCVOTNPAIETSSLPYSTLLRGSMLTALPVEASAPLIQAKLTLPLE
LGYRMESGITRRRDYILPPGAEIRLIQALSTARKRLPGLRSPESPV"
1470..4382
/gene="fhuA"
/note="XCC1749"
1470..4382
/gene="fhuA"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AA041040.1"
/db_xref="GI:21112842"
/translation="MTFSHSAAGTGAARARLRLHRSAMALISICVLLAARAYAOA
PLNAGSASVDDTVEYRGVRAALIKSOYIKRQASQIVDSVASEDIALPDRSVETL
QHVSVTIDRFMAKSPDHFSAEGSCVMIRGLQVRELNGRDFESAAGRSKFEDV
PAELMAGVYVKNPSAEIIEGLGVTNLRTRNPEPNRIGVANGVNDYVGMDSKKYK
PASLFLSDRMWTGIEGMFMIDIAHSELATSDIGOVPERDVAIAGSREY
YVPGVNRQOLFERTRDGIAAARFQKPSDAETIYKQFLRSREEMORARAFENDSN
NITTPAGTFTDIDERGAFKGAIVSSSMRGATGQVRFNTDNRSEGRVTTTDLV
GSHEFNMLQIKGMQVQSENEQDLTFVSATYLPGLYDVSQYPSVQIOLANPFT
QPSNTEFMAADHLGKNGRQSLSTFVLETFEDSSMLRFAFGMARADNOVANS
GNMGVISDNMAIPIGTNGLAMPDFTDOSLTFESFEGGVVPTTVLWVNSPL
VNVGASQMIQOIVLARGSNAPDTPOTPODROCEQOAAAYALXRENDENMGSLV
DENVGRVYVOTQAGRCQFQNLALANSPELQATYTCQYEDNNSQGYTNVLESLN
LRLRESDSLQMRFAASKAMARDYQLOPYVLLNTEINDAGATDFVAGNPMDEGT
VANOETALEWYEDSDMLYTLFVKKVYESTQREIYDNRPLVLRPYMMDGT
IKGAERGYPQFDSLPGMLNGFVANSNFTFVDSKGANATADPYTNTVYVGLPLEG
LSRRSYNLAGIYEKGPLSLRLAYNMRSRYLLTSSDAATRLPTMADDDYGLDGSFVRF
NDHLOGLQANNLNLTNVTKVMQPTSYEGGEVDPRLYTRSSFVNDRIYSLVLRMN"
4555..6747
/gene="citra"
/note="XCC1750"
4555..6747
/gene="citra"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AA041041.1"
/db_xref="GI:21112843"
/translation="MLPPOLGKPCMDYASTSPRLPPDDVAATTGGCVARAPAK
ISTCVLGSTSVSVORHRI PAGAVGORSIRERRSVMAISVATLLSAAYAKADATT
AAPASDATQOOLDYVOTGTRSSVKAQVQVQAEIVDSIAVEDIGKLPDNNVVAL
OHISGIOISRNQEGESSIAIRGLTVRTLENGDITFANDGRLSEDEVSALLEGVN
VKNPSADMIIEGLGTVLRTLRPLPFDYGRKTAGVQVNYNDIADLADGPARSGLESD
RHQGTGEGITLVNYSOOKSPRODPILEPYTOTDIPCYEGOSYVPHGACINNTI
GEREKRTGAFAMQWRPNDSVEYVTVLRSDDYFESWDDYFFALQINP10GLGDIQV
DNREFNVSQVNPESNTSLTERHSVTTDYISGAWMYNEKLLTSDFOYVADTTG
TRIVATGODTSQFQVNDPFRGLPRLSVTADASGAEGLADVNNYDQWRHYLDNKKDNK

gene
CDS
6853..7455
/gene="XCC1751"
6853..7455
/gene="XCC1751"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="TonB-dependent receptor"
/protein_id="AA041042.1"
/db_xref="GI:21112844"
/translation="MNYGLTFYKKYEGFIANA VFNENYVQGYQWQITRPVNGAKIRG
AELGYTFDFDLGWSLGGIQTNTYTVQSEAPSPATDINOSLTVPLEGLSKSYN
AILSETAFQGRVAVNMRSDMLVTYTAGGTGNLPYNNKGFQGLDASLRFNINDWMSI
SLDGVLDTLRRESYLGTESRDEYINDRRYGLTLRASL"
7567..9822
/gene="XCC1752"
7567..9822
/gene="XCC1752"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="AA041043.1"
/db_xref="GI:21112845"
/translation="MIAMIDPPKTRRTTGGAMAVLSLLLFSTPNYRAEPATSG
PYMRNSVAIGGGGFTVGLFHPAERGLAARDYGVGAQWMDAQQWATLQWLGAD
NLMGCTADAPADADATYLAAGTNRHARACNAALRSFNGRTERADLPKLGDN
QUGRAGERLAVDPHDGRVLLGSRDAGLRMSDRKARHAKYASFDALACATYARN
VGRQAVGIAFVGVDAASGNTCTTPPRITVGVSTESTSIVSEDAKSWAPACQPRG
LRPSHAGSDGHWYLSYGDQGPDLAMAGALMKFTPACGRNRRELSPIQPASGDFG
KCAVAVDPOOPQWYLSYTRRTPRDELRSYDQGNKMAPLADAFDHSAAFWTA
TPHMKMALIDEPDGNHAEFVYGGYIMASRNLODFAPORPLQMMFQDRLSETVLD
LISPMAGHLLSALGDIIDGRHDELDRLOLQVAPRITNGESIDAGAGAPQWVVRSGT
VDRNRNEIRALYSRGGQWRTAFASPEPPAGGAGSIALGADAQVYVAPBERGNNRT
SDGAWQWVDGIPDTAVVMAVDVDAKRRNTADVAGQLESTDAKSTRATGVQGS
PARDETRQOLRQEDPRAVAVVYLASPKGMVRMDGTLOVLSQPDARS LGIKALRA
GAPALYLAAGRYOGVGVFRSDDGQVOMORINDAHFRGPRPYSVTGDPRVAGRYEAT
GGGIRYVGPDR"
9926..11839
/gene="XCC1753"
9926..11839
/gene="XCC1753"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="sialic acid-specific 9-O-acetyltransferase"
/protein_id="AA041044.1"
/db_xref="GI:21112846"
/translation="MTPSHIRACILGVLSSGSAAPPSLPPLFADGAVLORDAMP
VVGMAKPGASTVSDGKQATKTAQAGQWKKLLPHTAGCPVSVSHDGCALQYRD
VLVDWVWLASGGSNEMFLAQSADGPOAAADADOLRRFKVPSKSVAPETRLTGGD
WKAAPENAGASVAVGFEFAKELRASTGVPIGIVDSTWGSALIEAMDAASGLLAQ
NOGALDAIKORDAQAQAAKRIGRWPKVDGOTPOARTEPFEDRDWDSIPVQOMELS
GYGMDGIMAYRTTFTLTAARAKAGVOLVGOIDSDIYVNDQOAGQTEKQMLPRV
VIVPAAALRAGVNOIARVVEDLSGGGGMGPDEORVVOIGSTORVALAGKFPAPQVRV
SLTDNKNOLPVLIXYNGMHIPLOPPPVKGVYVQGFNANPDGAVKRRDQFAMTQOVR
AERGARLPFLWYQOLANFRAGDGTGDSLMALIREOSQTTALPATGQAVITDTCPT
DIHPINKRQVGRHLAARHVAIVGETLVISAPFVADVAADQASVQFLOQASALQVR
GGGTGLOGFVAGADAKFRHFAQARIADRVIVSDAAYAAVAVRYGSENPADANVNR
EDLPVSFPTDWM"
BASE COUNT 2093 a 3908 c 3912 g 1997 t
ORIGIN
Query Match 4.6%; Score 116.8; DB 1; Length 11910;
Best Local Similarity 50.3%; Pred. No. 1.7e-08;
Matches 452; Conservative 0; Mismatches 427; Indels 19; Gaps 6;


```

gene
CDS
    /gene="XAC1771"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer/Genemark"
    /codon_start=1
    /transl_table=11
    /product="sialic acid-specific 9-O-acetyltransferase"
    /protein_id="AA036635.1"
    /db_xref="GI:21107969"
    /translation="MTASFVRGCLGLVLSPCAAVPTPLLLADGAVLORDOMP
    VMGSSPNAIAVSTGDKRATYKADATGOMKVRPAHAAGSPVPLVPODGGELQV
    VYGVWVLAGGSSNMEMPRLQAQSDGQAVAAADQOLROFKPKVMSVPOARLTGGE
    WKAAIPANAGETAVGVEFFAKELRASTGVPICIVNSTWGSALTEAMMDASLGLNDO
    NCGALEIQRDAAQAQATGKRIARMPKVEDMPQREAFDSDMSIPVKOMESS
    GYDGMGIMAYRTTITLSAAEKAGITLGGQIDSDTYVNGQVSGSTKEKNLPRV
    YQVPAALKAGVNHIAVVEDLSGGGGMGPDEORVOTSAQAKRALDQMKFPPAPAR
    VSLTDKNOLPTLLVNMTHLPLOPEPVKGVIVYCGETNATDTGAVKRYRQPAAMTROM
    RAERGCKTLPFLMNOIANFKAGDKELSPHALLRESOKTLALPATGCAVITIDENP
    TIHPNKRQVGHRLALAHVAVGETLVYSAVERKASFDGKAVLGGFDLQGSALQV
    RCGAVQGRFAGADQRFHPATAQIDGDVIVRSDAVAAVAVRGMSEBNDANLIN
    RALPYSPERTDW"
    /db_xref="GI:21107970"
    /gene="XAC1772"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer/Genemark"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AA036636.1"
    /db_xref="GI:21107970"
    /translation="MTRSLMSHMSPVARYIASSPARRRRLPRLLSALALAAACQSE
    QDDKAPPAVAVIAGSAAATPATPAAPAOETPMPOFVSKDKHALVNDAPFLILAO
    VNNSNYPGIMDKVMPAMQVGLGNTVOVPLAMEOVEBEGREDFSVFDLIRAREKH
    VRLVILMFATKRNNGPOYA PHVYKTDNQRPVRYVTEGSGISLSPHAQATLDDADKA
    FLAEMHKKVDPQHTVIMTOPENEAGITGVSFDESPMAOKLFDGOVDELLIKRMKQ
    PGTMAOVFGADADEIETHAVSIGRTYIDQIAQGAELPLPMYVNAALRGFPNPOQOY
    ASGGPNDLDVWKAAGPAHIDLAPDIYMEPYMTYLERVAPRDNALFVLETGNRP
    EXARYLPTLGNHIGMSAFGIDYTRYSNYPLGAKHNEETLAPFALGQAVASMGAR
    PKAASEGKLGHTAEERGOAVOELKIDRMSATITYGVOFMFKGEPNPRPTGAL
    IAEISDFELVYGHARTLTHASDATAMHITIDRVEGVYEGDNKRFQNMNDDQDY
    GYRFSVPOLKIKATYK"
    /db_xref="GI:21107971"
    /gene="XAC1773"
    /note="XAC1773"
    /gene="XAC1773"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer/Genemark"
    /codon_start=1
    /transl_table=11
    /product="alpha-xyloridase"
    /protein_id="AA036637.1"
    /db_xref="GI:21107971"
    /translation="MEIPAOLEWRPDLRLRQFORAAAQDQAGNLQISAPATPRRR
    RGIASPPYCAAPAPRPRSKGVIQWOTIRSGAPARARSLSLAQSLSLMLANA
    AHAQEVKCAADVTVPASGAAPVRLQVDAIGIIVASADDDFARSLSLMLANA
    DTAQVABQDGSVOLTKGVATSIITVDHSPADNGPVSLEYSEVSGSFALANGL
    KOYLSYROROSPDDEALYGFQHOOGMNOGRNELDONNIDMAVPLVYASRNLG
    LMDNSITRLGPRGQIPKPTITLIDAKGKAGALYARAIKNGHILERRESVNYOY
    LSDLTKPKATITKDNKSNQYTWEGEIALTGEETSLYSEVAKLIVDGLVNDRA
    WQNMNPMHNERKLDEPGQRHTVYEMVLIDPSYIALHRLPFLPAEKDLSTLMEBA
    GMDIDYFVSADYQAVAGYRALTGKSTMLPKMAAGFQMSRERYSOBELIVASVLEY
    RKRKSLIDNIVLDMSYWPEANAGSHDFDQHPDPGMKAVYDMIAOILSIMPYR
    PTTANYKEIDAGFMFKRVYEGELDMICKYKXNSYDYSEAKQALYKROTKENKNS
    KGFDAWMDADEPVAHNSLIDIERKARTTPNALGSSTETFNSTPLDHTGAVYGDRAA
    DDKRYFTLSRKGAGTQRNAVAVMSGDIVSRMDMDQDISGVMALSGLPMWTFDIG

```

```

gene
CDS
    /gene="XAC1774"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer/Genemark"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AA036638.1"
    /db_xref="GI:21107972"
    /translation="MTRRELKATGAIAVLGAVPAAQAQATQASKTALPYVAAA
    LQILWREPNEWEALVGNRGRIAMWGMGIAHERLQNEEDLYAGPYDSTSPALA
    ALPOVALTFAGRYAEKADAKLSRPLKMPYQPLGDLDFRADGIDSRYROL
    DIDTAVATTFPRSGAVHREYFVCAOACIVYRLSCDRGGISLRVGLIDSPQTEV
    ARGGILFSGRNSFAGITGRPLRPLRVYPOVSGKLSQVPRRLRIDADVYLLISA
    ATSYORFADVDDPLALYARAKKALDPLALRHLHDHQRFRRAVIDGSEAV
    QLPTEBRYORFAEGNDPALALYHGYRLLLCSSRHPQANLQIMNDLQPWES
    KYTININIMNWPSEANALHCEVEPLEAMLDLAQGTHTARAIYDADGAVVHNNTD
    LMRQAGPIDGAOMSLMPMGVNLLOOLMDRMVDGRDALYSKYVPLFKGAEEFVNL
    MRDPOTGAVTNPSMSPNDHPGAVVCGVPSMDLOLDLDFACQIAMSLLGIDQOL
    AQAALAREQOLPPNRIGKAGQLOENQDMDQAPETNNHVVSHLYALHPSQINLRDT
    PELAAARRSLEIRGDNATMGIGSNGLNIMARLADENHARTLOLISERTYPNLFD
    AHPFQIDNFGGTAGITETMLQSGWSPFLPLALPKAPRSGSVRLRVGASVLDLE
    WEGRLQQAARLSDRGRYQLSYAGOTLDELGAGRTQOYGLNNRLVYQ"
    /db_xref="GI:21107973"
    /gene="XAC1775"
    /note="XAC1775"
    /note="identified by sequence similarity; putative; ORF
    located using Blastx/Glimmer/Genemark"
    /codon_start=1
    /transl_table=11
    /product="D-xylulokinase"
    /protein_id="AA036639.1"
    /db_xref="GI:21107973"
    /translation="MSLVGLDVGTVSVKLVAVYDPODRAVVAITAAPMLISRDGTR
    EQQAQWIMIGIYHCFQDLGEDBARVGSVSQOQGFVPVADGVTAPVVKMCDTS
    TLECEIEMAAVGAAGSVAAGNPIIMAGYTASKLPLWTKRHRDADVAAMTYVLPDHY

```

Query Match 4.28; Score 105.4; DB 1; Length 14520;
 Best Local Similarity 48.38; Pired No. 8.8e-07;
 Matches 487; Conservative 0; Mismatches 501; Indels 21; Gaps 6;

```

OY 72 GAAGCTCAACCTCGGCGGCGGCGGCTGTCGCCCGGATCATCTTCATCCCAAGAC 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 GCACAGCGTGGCCATCGGGGGGGCGGCTTTCACCGGCTGCTGTTTATCCCGCGA 292
OY 132 AAAAGCGGTAGCATATGACGAACAGATATGCGGGCTATACCGCTCAACGCCGACA 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 ACGGATCTTGCCATGACGACGACCGATGCGGAGTGCCTATGCGTGGGACGGCGAGC 352
OY 192 CTCATGACCGCGCTGACGAGATGGATGCTGATATGCCGGCTGGGACAACTGGGCAT 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 GCACGACTGGGCTCGCTGACGATGTTGGGCGCGGAGATTGGATGTATGGGAAT 412
OY 252 CGAGCGTGTGCGCTGATCGGACGACATCAAAAGGTATGATCCGACATCGCATGTA 311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 413 CGAGCGCTTTGCGGTGATCTCGCGACCGAAGCGCTTGTACTGCTGCTGCGACACTA 472
OY 312 TACGACAGCTGGGATCGAGTAATGAGGCATCTATTCGCTGACAGCCGCGCGAAC 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 473 CATGACAGAACGGG---CCGCAATGCGCGGCTGCTGCTGCTGATGCGCGTAAGCA 529
OY 372 GTGCTCTTACCAACTTGGCTTCAAGTGGGGGATACATGCGACAGACGGGAGCGG 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 530 GTTCAAGCGCGGACCTGCGCTTCAAGTGGGACCAATCTTGGCGCGCGCAATGG 589

```


[illegible]

REFERENCE	2. (bases 1 to 16150)
AUTHORS	Parkhill,J.
TITLE	Direct Submission
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk On Jun 27, 1998 this sequence version replaced gi:2916933. Notes: Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/projects/M.tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of TB genes implemented in Tbpase (Krogh), supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. Location/Qualifiers source 1. 16150 /organism="Mycobacterium tuberculosis H37Rv" /strain="H37Rv" /db_xref="taxon:83332" 14410..>16150 /organism="Mycobacterium tuberculosis H37Rv" /strain="H37Rv" /db_xref="taxon:83332" /clone="I237" <1. 14510 /note="fragment designated v044. Does not represent a physical clone" 157..162 /note="possible RBS for Rv0977" 171..2942 /gene="PE_PGRS" 171..2942 /gene="PP_PGRS" /note="RV0977, (MTV044.05), len: 923. Unknown glycine richprotein, similar to other PGRS-type sequences e.g. gpIAL0091 MTV004_1M, tuberculosis sequence (1125 aa), 45.4% identity in 959 aa overlap: or gpI280225 MTVCY44.1 M. tuberculosiscosmid (778 aa), 51.5% identity in 750 aa overlap. TBPasescore 150.868" /codon_start=1 /transl_table=11 /product="PE_PGRS" /protein_id="CAAI1576.1" /db_xref="GI:2916938" /db_xref="SPREMBL:O53890" /translation="MSFVVTAAPVLAASAADIGASISSEANMAAVRTALAPAA DEVSATIALPSTYARDYOTLSQVTAHFVRQITLTNAGOLIAVDGNCLELTED QVLEGINAPQTLYGRPLIGDTHGAFTGGONGAGGLIMNGNGSGSAGPAQPGRG GDJHGSGHGHGVGGPGIAGAAGTAGLPFGANGANGSCGTICAGACGANGLLFNG GAGCGGSGSLIGSGSGTGAGMAAGPAGCTGIGTIGGIGAGVGAGSALFEHGIG I NKGWGGGOGGAGGAGNGMAEBGITVIGEORGOGGDDGAGAGATGGSAGIGSSO GAGCHGGDGOGGAGSGGCVGGGAGAGDAGCIGTCGCNGSITGCAANGCNGRC GACGMAATAGSDGNGGNGGNGEYVGSAGAGCTGDGAGAGAGAPHPFOOPAPO GLPIGTGTGEGGAGAGGAGDGGQIDIGFDGRGDDGPGGGGAGGSGSTFNAAQN NGDDGAGVGAGAGTGTGTGAGADGRCGDDGRCGDDGMACHGAAPFSRGAYAGE GSGGAGGAGNAGAGTGTAGSGAGGFGNGADGNGCNGNGCGGINTEGNTAGAG GTGGGLGTLGGHNIGLNATAGTGGTGTLTNAATVPLQVNTTEPVVEISLNGGOMP VILDRGSTCLNDOSPTLRONRGPVICTGACAGCTTYNNYTSSTVPFCGLTIPT SUVAVYSSSPGTLGNFLRSAGAVGULGIPNNGPOTSIYTAMPGLLNNGCYLIDES A GILOEPNLTLGGITTSI GAPISTVAQDINDGLQAAPVAFDSGGINGTIPSALSLPS GCFFVPGATTIVSYTDGDTLLSYTTTANPFVTSGGVMNTGHVPFAQDPITYVSP TAIGTTTTN"
FEATURES	
source	
source	
misc_feature	
RBS	
gene	
CDS	
gene	complement(3159..4154) /gene="PE_PGRS" complement(3159..4154) /gene="PE_PGRS" /note="RV0978c, (MTV044.06c), len: 331. Member of M. tuberculosis gly-, ala-, asn-rich protein family, pgys family,similar to e.g. gpI295387 MTCYA10_19 M. tuberculosis cosmids (461 aa).73.6% identity in 277 aa overlapscore is 0.861" /codon_start=1 /transl_table=11 /product="PE_PGRS" /protein_id="CAAI1577.1" /db_xref="GI:2916939" /db_xref="SPREMBL:O53891" /translation="MSFVNAPQLYSTAADARIGSINTANTAAATTOVLAADD EVSATIALFGSHGOHYOAISAQVAAVOORVLASQGSTYVAEAASATPLDNVID ALNAPQSITGRPLIGDGANGLDGTGOAGNGMWNGNGSGAGAPQAGACGAGG LIENGAGCTGAGVASLARAGTAGAGRPVGGIGAGCGVAGCAGAACVTTITHASFNR DPHGVAVNGMVNYTNFSGTSVSNINATPNTVTSPTTINCSGSVASVPTLVTV TNDSTNVSVIPPTNTPMGSPITGTAPGVAVNPGEVYVTFNPAQDYVSIS" complement(4160..4166) /note="possible RBS for Rv0978c" complement(4468..4794) /gene="RV0979c" complement(4468..4794) /gene="RV0979c" /note="RV0979c, (MTV044.07c), len: 108. Unknown." /codon_start=1 /transl_table=11 /product="hypothetical protein Rv0979c" /protein_id="CAAI1578.1" /db_xref="GI:2916940" /db_xref="SPREMBL:O53892" /translation="MGVTDGHTDGLGGELPLRAPASCIRARLSALGHSNGSPSG MCFRRVGAATTASTYTHRIPIPEDGPAQFRAGVGRRODTVVAPMVYGLMDNRRP GWQMPs" complement(4801..4804) /note="possible RBS for Rv0979c" complement(4876..6249) /gene="PE_PGRS" complement(4876..6249) /gene="PE_PGRS" /note="RV0980c, (MTV044.08c), len: 457. Member of M. tuberculosis gly-, ala-, asn-rich protein family, pgys family. similar to many e.g. gpI295387 MTCYA10_19 M. tuberculosis cosmids (461 aa), 66.7% identity in 405 aa overlap: or gpI295844 MTVCY43_2M. tuberculosis cosmids (741 aa). 53.0% identity in 394 aaoverlap" /codon_start=1 /transl_table=11 /product="PE_PGRS" /protein_id="CAAI1579.1" /db_xref="GI:2916941" /db_xref="SPREMBL:O53893" /translation="MSFVVAAPQISTAAADAARTGAINANTAAATTOVLAAND EVNSTIALFGSHGOHYOAISAQVATDERVYLASQSTTYVAEAASATPLDNVID ALNAPQOSTLGGRPLIGDGANGLDGTGOAGNGMWNGNGSGAGAPQAGACGAGG LIENGAGGAGGAGQGFPEFAGANGDGTGTGAVASLARAGTAGAGGAPAGTIFVAGGVTGGVAGHGLIGVGNGHGGDGTGTGAVASLARAGTAGAGGAGGAGVAGGAGAGVATITIHASFNRDPHGVAVNGMVNYTNFSGTSVSNINATPNTVTSPTTINCSGSVASVPTLVTV TNDSTNVSVIPPTNTPMGSPITGTAPGVAVNPGEVYVTFNPAQDYVSIS" complement(6255..6261) /note="possible RBS for Rv0980c" 6614..7306 /gene="RV0981" 6614..7306 /gene="RV0981" /note="RV0981, (MTV044.09), len: 230. Unknown, but highlysimilar to various regulatory proteins of two-component response regulators e.g. BJY09666.1

B.japonicum fragA, fragB and rpoH3 (226 aa), 43.8% identity in 224 aa overlap; or BSAJ2571.44 B.subtilis 168 56 kb DNA (228 aa), 46.4% identity in 224 aa overlap; or MFCY31.34 M.tuberculosis cosmid (236 aa), 50.7% identity in 225 aa overlap; or MCB57_27 M.leptaeosmid B57 (233 aa), 50.2% identity in 225 aa overlap. Contains PS00217 Sugar transport proteins signature 2^a

/codon_start=1
/transl_table=11
/product="hypothetical protein RV0981"

/protein_id="CAI17580.1"

/db_xref="GI:2916942"

/db_xref="SPTREMBL:O53894"

/translation="MSVRLIVDDRAVRESLRSLSFNGSYVELADHVEALDMIAS DRPALVDYMMPLDGLVEVROLRGVDDLPILVLRDSVSRVAGLAGADYLP KPELELLARVALLRRTKPEDAASMAHRSFLTDLPYRVRNRRORSUTRTER ALEMLTANRRVLRSLRILEWGFDPISGNLEVYGLTKRTEADGPRLIHTV RGVGYLRRTTP"

/gene="RV0981"

/note="PS00217 Sugar transport proteins signature 2"

/gene="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

/note="RV0982"

Query Match 3.7%; Score 92.2; DB 1; Length 16150;
Best Local Similarity 43.9%; Pred. No. 8.8e-05;
Matches 443; Conservative 0; Mismatches 563; Indels 3; Gaps 1;

1930 GACGCGCATATTCGCGTCCACAGACTCGGACAGACCTTTGGCCAGTCTCC---ACC 1986
1079 GACGCGCATATTCGCGTCCACAGACTCGGACAGACCTTTGGCCAGTCTCC---ACC 1020
1987 GCGCTGACCAACACCTTACAGATCGCCCTGGGTGTGGCTCAGGCTCAGTGAACCTG 2046
1019 GCGGCTACCGCATCGCGCTGATTCGCCCGTGGCGGAACAACGCGGAGCGTGGCCGCC 960
2047 TATGCTTTCGCGCGCGCGCTGAGGAGCTCGCTTACCGCAGTGTGAGACAGCGCGCC 2106
959 GACCCCGCGCGCGCGCGATGCGCGCCGATGCCCGCGGATGCGCGCGCGCGCGG 900
2107 TCCTGACGAGCATATTCGCGTCCACAGACTCGGACAGACCTTTGGCCAGTCTCC---ACC 2166
899 ACCGCGACCATGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
2167 GCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
839 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
2227 GCTCAGGAGACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2286
779 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
2287 AGCAGTACCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2335
719 AGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 671

RESULT 14

AE006985/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

gene

gene

gene

gene

gene

gene

16168 bp DNA linear BCT 27-APR-2001
Mycobacterium tuberculosis CDC1551, section 71 of 280 of the complete genome.
AE006985.1 GI:13880571
Mycobacterium tuberculosis CDC1551.
Mycobacterium tuberculosis CDC1551.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 16168)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., Deboy, R., Dodson, R., Gilm, M., Haft, D., Hickey, E., Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uitterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
Unpublished
2 (bases 1 to 16168)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., Deboy, R., Dodson, R., Gilm, M., Haft, D., Hickey, E., Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uitterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1..16168
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="Clinical strain"
115..2918
/gene="MT1004"
/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:U123456; identified by sequence similarity; putative; PE_PGRS family protein, authentic frameshift."

```
gene complement(3135. .4142)
CDS /gene="MT1006.1"
      complement(3135. .4142)
      /note="MT1006.1"
      /note="similar to GB:AL123456; identified by sequence
      similarity; putative"
      /transl_table=1
      /codon_start=1
      /product="PE_PGRS family protein"
      /protein_id="AAK45254.1"
      /db_xref="GI:13880572"
      /translation="MSFVVAHPOLVSTAADARIGSAINTNTAAATTOVLAANO
      EVSTAIALFGSHGHOYAISAQVAAYQOREVLALSOAGSTVAVAEASATPLOQIEQ
      ALLGVINTPEALVGRKLGIDGAGAPGTGAGGILMGNGNGSGAPGAGAG
      GAGGLGNGAGAGTGAVALRAGTAGAGRPGGIGAGCGAGAGCAAVTTTH
      ASENDPVAVNPGRGNVYTNFGSGVSYINPATMTMTTIGNGSPGAVSPVIG
      LVFVTFNPSNTSVSDPTTNTVTGSPITVGTAPGVAVNPVGEVYVNFADPTSVI
      S"
gene 4321. .4845
CDS /gene="MT1007"
      4321. .4845
      /gene="MT1007"
      /note="identified by Glimmer2; putative"
      /codon_start=1
      /transl_table=11
      /product="hypothetical protein"
      /protein_id="AAK45255.1"
      /db_xref="GI:13880573"
      /translation="MSAGETRADENSQVQKHSLLTLNGQRTPEAPCRPRVRRPLANS
      ASTRPLPMSPIPVPPPHHRCNDRELVTTPGPNPGPELGRVFNRPYRRCRRCR
      LKARLAEHTKGRPTMAVPKRRKRSNTRRSQMKAKTELVTGTVAGHAKVPRRL
      LKAARLGLIDFKR"
gene complement(4864. .6294)
CDS /gene="MT1008"
      complement(4864. .6294)
      /gene="MT1008"
      /note="similar to GB:AL123456; identified by sequence
      similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="PE_PGRS family protein"
      /protein_id="AAK45256.1"
      /db_xref="GI:13880574"
      /translation="MGRPLNTVAVRISMFSFVNAPOLVSTAADARIGSAINTNTAN
      TAAATTOVLAADHDEVSTAIALFGSHGHOYAISAQVAAYQOREVLALSOAGSTVA
      VEAASATPLOQIEQALLGVINTPEALVGRKLGIDRAHAGRTGOAGAGGILMGNG
      GNGSGAGGAGAGAGAGLIGNGAGCGGCGGLPPAGANGAGAGAGGMLFNGGAG
      GYGAGGAGATTTCVAGCGDGTGCGVGHGGLISVGCHGDCGTGTGCAVSLARAGTAG
      GAGGPGAGTGGAGGAGCAAVTTTHASFNDPBGVAVNPGRGNVYTNFGSNTV
      SVIDPVTNVTGSLTDGNGPSGVAAPVGLFVTFNPSNTSVSDPTTNTVTGSPITV
      GTGAYGVAVNPGRGNVYTNFGSNTVSDPTTNTVTGSPITVGLDPTGVAVNPVTVGV
      YTTNSLDPTVSIVTGEPARSVCSAAI"
gene 6614. .7306
CDS /gene="MT1009"
      6614. .7306
      /gene="MT1009"
      /note="similar to PID:1465701 PID:1857994 GB:AL123456;
      identified by sequence similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="DNA-binding response regulator"
      /protein_id="AAK45257.1"
      /db_xref="GI:13880575"
      /translation="MSVRILVDDRAVRESLRSLSFNGYSVELAHQVEALDMIAS
      DRPALVDVMMPRDLGLEVCROLRGTDGLPIVLTARDYSERVAVNGCRRLSUTTER
      KPALEELLARRALRRKEDAEASAMRSDTLDPVTRVNGCRRLSUTTER
      ALLEMLNPRVLTFRSLRLEWGFDEPTSGNALEVVYGLRRTEADGEPRLLHTV
      RGVGLRETPP"
gene 7306. .8820
CDS /gene="MT1010"
      7306. .8820
      /gene="MT1010"
      /note="similar to GP:4455704; identified by sequence
```

```
similarity; putative"
      /transl_table=11
      /product="sensor histidine kinase"
      /protein_id="AAK45258.1"
      /db_xref="GI:13880576"
      /translation="MMWFRRRDRAPLRATSSLSLRMYLAMSVMVAVVLMSPAVY
      AVISAALYSDINQSRALDASGLAADPGKAIETGTAAYSDVAAVNPQOSIYTA
      OOPGOTLPVGAERKAVIRGELFMSRRTADQVLAIRFNGSSLLISKLPTFAVMN
      KLRWVLTIVGGVAVVAVAGMVRAGLRTPVGRLTTEAERVARPDLDLRPIPVGSD
      LARLTAFRNLRLALAESRERARLVTPAGIELRPLTSLRNVELLASNAPGAPRL
      PKQENVLDLRADLAQIEELSTLVGLVDLSRQDAEYVHEPVDMAADVDRSLERRR
      RNDILPDEVIGMOYGDYAGLSRMLNMDNAKMSPPGCVGRVLSQLASHAELV
      VSDRPGPIVVERRLVPERFYASARALPGSGGLAIIVKQVVLNHHGGLLRREDTDPG
      GQPPSTSYVLLPGRMRPIPLPGATAGARSTDIENSGSANVISVESQSTRAT"
gene 8918. .10258
CDS /gene="MT1011"
      8918. .10258
      /gene="MT1011"
      /note="similar to GP:4455705; identified by sequence
      similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="heat shock protein HtrA, putative"
      /protein_id="AAK45259.1"
      /db_xref="GI:13880577"
      /translation="MTNHPRYSPPOPGTPEGVAGGQOQTSQOQDMRYPPSPPOPT
      ASYVFNPNAPGSPGPPVAAASAPSIPIANMPGSGVEQAAKVPVSVWLETDLCROS
      BEGSGITILLASBLLITNNHVIATAAKPLDGPPTVTFSDGRTAPPTVAGADTSD
      IAVNVQGVSGLTPISSSDLRVGOEVLVLAGSLGLEGTGTGIVSALNRVSTTG
      BAGNONTVIDAIDTDAIINPGNSGALNMNAQIVGNSAATLQADSDAOSGSLG
      GVAIPVQDAKRIADELISTGKASHASLGVOYNDGTGAKIVEVACGAANAAPVK
      GVYVTKVDPRPINSADALVAVRSKAPCATVALTFQDPSGCSRTVQVTLKAEO"
gene 10255. .10803
CDS /gene="MT1012"
      10255. .10803
      /gene="MT1012"
      /note="similar to GB:AE000511 PID:2313931; identified by
      sequence similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="molybdopterin biosynthesis protein"
      /protein_id="AAK45260.1"
      /db_xref="GI:13880578"
      /translation="MMKVAOCSKIGYTVAPMEORAEVLVGRALVVVDDPRANGDED
      HSGPIVTELPENAGVNVGVAVSADEVEIRALMTAVTIGVDLYVSGGVPRDY
      TPEATRDLDREILGIAEIRASGLSAGIVDAGSLGAGVSGTLVNVNLAGSRVAVR
      DGMATLNPLAAQITIGQLSSLEI"
gene complement(10823. .11278)
CDS /gene="MT1013"
      complement(10823. .11278)
      /gene="MT1013"
      /note="similar to GB:L29458 SP:P23867 GB:U08371 GB:X52114
      PID:43138; identified by sequence similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="large conductance mechanosensitive channel"
      /protein_id="AAK45261.1"
      /db_xref="GI:13880579"
      /translation="MLKGFKEFLARGNIVDLAAVAVIGTAFALVTKFTSDITPPLIN
      RIGVNAOSDVGLIRIGIGCOTIDLVLSAAINFLFAVYFLVAVLPYMLRKKE
      VEOPGDTQVLLTEIRDLAQTNDSGPRHGRGPRSPPTDGPRASTESQ"
gene 11601. .12347
CDS /gene="MT1014"
      11601. .12347
      /gene="MT1014"
      /note="similar to GB:AE000782; identified by sequence
      similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="ABC transporter, ATP-binding protein"
      /protein_id="AAK45262.1"
```

```

/db_xref="GI:1380580"
/translation="MNRPIVQLSNLSMTFREGERRQVLHITFEPGEFVALLQ
SGSGKSLINLSIGIEPTGDTGKNGRITKPTLRDLRFDIGIVFOFNLIPT
LTVENITLPOELAGVSRKAAYVADLLEFVGMADREPTPKLSSGEODRVAISR
LAHNMLVLADPEPTGNLSDSDYKLVLDLITRQAGTLIMATHSPSMTOHADRYN
LQGGFLPVAIVRENQTDQPASTILPISYE"
gene
12340..14907
/feature="MT1015"
12340..14907
/feature="MT1015"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11

Query Match      3.7%: Score 92.2; DB 1; Length 16168;
Best Local Similarity 43.9%: Pred. No. 8.7e-05;
Matches 443; Conservative 0; Mismatches 563; Indels 3; Gaps 1;

1330 GACGGCAGTGAAGAAATTCCTGCTCAGAGACTGGCTTCGACACCGCGGAAGCAAGCTA 1389
1653 GAAGGTCGCGCCGCTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1594
1390 TTGGCGGAGTGGAGACACACAGCGCTTCTGCTTGCAGAGAAACAGCACTCGGAGACA 1449
1593 CCGGCGCGCATGAAGCCGATGTGCTGCGCCGCGCTGCTCAACCGCGCGCGCGCGCGCG 1534
1450 TCGCGCGACAGCGTCTGGGCAACGCGCACATGAGGCGCGCATGCGAGCGCTGCACTACGCG 1509
1533 GCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1474
1510 GGGAACTCGGTCAAGACGCTGCTGCGCGTGGCAACACCGCGCGCGCGCGCGCGCGCGCG 1569
1473 GTAGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1414
1570 ATCTGCTCGAGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1629
1413 GCGCCCTCGGCGCGCTGCCAACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1354
1630 GCGCGCAGCGTGGCGCTATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1689
1353 ATCACTTCCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1294
1690 GCGGTGCAACGCTCGCAGTTCCAGGGCAGCTTTCCTCGCTCGAGCGCTGCCCGCGCG 1749
1293 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1234
1750 GCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1809
1233 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1174
1810 TTATACGTACGACAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1869
1173 GCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1114
1870 GGGAGCATCGGGGATATGCTGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1929
1113 GATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1054
1930 GAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1986
1053 GACCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 994
1987 GCGCTGACCAACAGCACTTACGAGTCCGCTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT 2046
993 GCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 934
2047 TATGCTTGGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2106
933 GACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 874
2107 TCTGAGACGAGATTCAGAGGCTCCGAGGGCTTCGAGTCCATGACAGCAAGCAAGTTCGCC 2166
873 ACCGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 814

```

```

QY 2167 GGCAGCGCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2226
DB 813 CTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 754
QY 2227 GCTCAGGACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2286
DB 753 GCGCGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 694
QY 2287 AACAGTACCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2335
DB 693 ACGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645

RESULT 15
AL161725
LOCUS
DEFINITION
AL161725 181179 bp DNA linear PRI 16-JAN-2001
Human DNA sequence from clone RP11-165F24 on chromosome 9. Contains
the 3' end of the gene for a novel protein (similar to Drosophila
CG6630 and CG11376, KIAA1058, rat TRC), an RPL12 (60S ribosomal
protein L12) pseudogene, ESTs, STSs, GSSs and a CpG island,
complete sequence.
ACCESSION
AL161725
VERSION
AL161725.13 GI:10045359
KEYWORDS
HTG; CpG island; KIAA1058; RPL12; TRC.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 181179)
REFERENCE
Laird, G.
Direct Submission
Submitted (11-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Sep 9, 2000 this sequence version replaced gi:3864498.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/projects/C-elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unseq' feature key.
RP11-165F24 is from the library RPL11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-165F24. The true
right end of clone RP11-5906 is at 17684 in this sequence. The true
left end of clone RP11-31F19 is at 158244 in this sequence.
FEATURES
source
1..181179
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-165F24"
/clone.11b="RPL11.1"
/misc_feature
9..757
/feature="match: GSS: Em:A0417012"
misc_feature
22..474

```

```

misc_feature
/note="match: GSS: Em: A0318836"
56..480
/note="match: GSS: Em: A0390327"
809..1072
/note="MIR repeat: matches 8..264 of consensus"
1121..1289
/note="t2 repeat: matches 2590..2750 of consensus"
1881..99118
/gene="ba165f24.1"
mRNA
join<1881..1998,4093..4163,5291..5429,6048..6149,
10073..10168,10840..11074,13634..13795,16376..16548,
20194..20289,24334..24429,30648..30797,33009..33122,
38781..38936,40793..40932,48645..48814,51931..52070,
54264..54446,54812..54941,55911..55998,60748..60844,
62225..62359,63565..63717,66029..66187,67738..67838,
68646..68838,73108..73251,75149..75280,75738..75872,
77290..77379,80233..80469,83647..83790,85874..85980,
97380..97550,98022..99118)
/gene="ba165f24.1"
/product="ba165f24.1.1 (novel protein similar to
Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
1))"
/note="match: CDNAS: Em: AB037816
match: ESTs: Em: AW504701 Em: AW500066 Em: AW500946 Em: H89666
Em: W30966 Em: AW499621 Em: A1524110 Em: AJ394589 Em: AA146391
Em: AW461668 Em: AW98225 Em: AW504408 Em: AW337875
Em: A1927976 Em: AW500913 Em: AA374756 Em: T63639 Em: AA321435
Em: AA02794 Em: AA212554 Em: AA146408 Em: AA365573 Em: N36754
Em: C01477 Em: C01637 Em: T80502 Em: AA11026 Em: A1192903
Em: AA170814 Em: A1290154 Em: AA251652 Em: AA007697
Em: A1660115 Em: AA893933 Em: A1422078 Em: AA552743 Em: N48396
Em: AA401587 Em: N30451 Em: N57465 Em: AA401548 Em: A1862755
Em: AA400038 Em: AW511478 Em: AW511478 Em: AW207838 Em: A190239
Em: AA653937 Em: A1693055 Em: AA213632 Em: A170724 Em: A158119
Em: N30504 Em: A1480327 Em: N30464 Em: A1769537 Em: AA742950
Em: A1621032 Em: AA251553"
/evidence=not_experimental
join<1881..1998,4093..4163,5291..5429,6048..6149,
10073..10168,10840..11074,13634..13795,16376..16548,
20194..20289,24334..24429,30648..30797,33009..33122,
38781..38936,40793..40932,48645..48814,51931..52070,
54264..54446,54812..54941,55911..55998,60748..60844,
62225..62359,63565..63717,66029..66187,67738..67838,
68646..68838,73108..73251,75149..75280,75738..75872,
77290..77379,80233..80469,83647..83790,85874..85980,
97380..97550,98936..99118)
/gene="ba165f24.1"
/product="ba165f24.1.2 (novel protein similar to
Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
2))"
/note="match: ESTs: Em: AW337875 Em: C01477"
/evidence=not_experimental
join<1881..1998,4093..4163,5291..5429,6048..6149,
10073..10168,10840..11074,13634..13795,16376..16548,
20194..20289,24334..24429,30648..30797,33009..33122,
38781..38936,40793..40932,48645..48814,51931..52070,
54264..54446,54812..54941,55911..55998,60748..60844,
62225..62359,63565..63717,66029..66187,67738..67838,
68646..68838,73108..73251,75149..75280,75738..75872,
77290..77379,80233..80469,83647..83790,85874..85980,
97380..97550,98936..99002)
/gene="ba165f24.1"
/codon_start=2
/evidence=not_experimental
/product="ba165f24.1.2 (novel protein similar to
Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
2))"
/protein_id="CAC22149.1"
/db_xref="GI:12274836"
/translation="MLVYVQRLNFVNKLASARNITIKIOFMCEDASNAMPVIEGK
SSGPELOEVYAVYHNKSPDEYEVKIKPAKIVVHNLFTYHHSICOOKGASV
ETLLGSMLPTLNERLOTGSCYCLPALEKLPKPNHSMASAEVPLQNPPIKRAEENK
VFNIEVQAVSSVHTODNHLKEFTLCHSLESQVTPPIRVLDOKISEMALEHKLSTII

```

```

misc_feature
/note="match: GSS: Em: A0318836"
56..480
/note="match: GSS: Em: A0390327"
809..1072
/note="MIR repeat: matches 8..264 of consensus"
1121..1289
/note="t2 repeat: matches 2590..2750 of consensus"
1881..99118
/gene="ba165f24.1"
mRNA
join<1881..1998,4093..4163,5291..5429,6048..6149,
10073..10168,10840..11074,13634..13795,16376..16548,
20194..20289,24334..24429,30648..30797,33009..33122,
38781..38936,40793..40932,48645..48814,51931..52070,
54264..54446,54812..54941,55911..55998,60748..60844,
62225..62359,63565..63717,66029..66187,67738..67838,
68646..68838,73108..73251,75149..75280,75738..75872,
77290..77379,80233..80469,83647..83790,85874..85980,
97380..97550,98022..99118)
/gene="ba165f24.1"
/product="ba165f24.1.1 (novel protein similar to
Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
1))"
/note="match: CDNAS: Em: AB037816
match: ESTs: Em: AW504701 Em: AW500066 Em: AW500946 Em: H89666
Em: W30966 Em: AW499621 Em: A1524110 Em: AJ394589 Em: AA146391
Em: AW461668 Em: AW98225 Em: AW504408 Em: AW337875
Em: A1927976 Em: AW500913 Em: AA374756 Em: T63639 Em: AA321435
Em: AA02794 Em: AA212554 Em: AA146408 Em: AA365573 Em: N36754
Em: C01477 Em: C01637 Em: T80502 Em: AA11026 Em: A1192903
Em: AA170814 Em: A1290154 Em: AA251652 Em: AA007697
Em: A1660115 Em: AA893933 Em: A1422078 Em: AA552743 Em: N48396
Em: AA401587 Em: N30451 Em: N57465 Em: AA401548 Em: A1862755
Em: AA400038 Em: AW511478 Em: AW511478 Em: AW207838 Em: A190239
Em: AA653937 Em: A1693055 Em: AA213632 Em: A170724 Em: A158119
Em: N30504 Em: A1480327 Em: N30464 Em: A1769537 Em: AA742950
Em: A1621032 Em: AA251553"
/evidence=not_experimental
join<1881..1998,4093..4163,5291..5429,6048..6149,
10073..10168,10840..11074,13634..13795,16376..16548,
20194..20289,24334..24429,30648..30797,33009..33122,
38781..38936,40793..40932,48645..48814,51931..52070,
54264..54446,54812..54941,55911..55998,60748..60844,
62225..62359,63565..63717,66029..66187,67738..67838,
68646..68838,73108..73251,75149..75280,75738..75872,
77290..77379,80233..80469,83647..83790,85874..85980,
97380..97550,98022..99118)
/gene="ba165f24.1"
/product="ba165f24.1.2 (novel protein similar to
Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
2))"
/note="match: ESTs: Em: AW337875 Em: C01477"
/evidence=not_experimental
join<1881..1998,4093..4163,5291..5429,6048..6149,
10073..10168,10840..11074,13634..13795,16376..16548,
20194..20289,24334..24429,30648..30797,33009..33122,
38781..38936,40793..40932,48645..48814,51931..52070,
54264..54446,54812..54941,55911..55998,60748..60844,
62225..62359,63565..63717,66029..66187,67738..67838,
68646..68838,73108..73251,75149..75280,75738..75872,
77290..77379,80233..80469,83647..83790,85874..85980,
97380..97550,98936..99002)
/gene="ba165f24.1"
/codon_start=2
/evidence=not_experimental
/product="ba165f24.1.2 (novel protein similar to
Drosophila CG6630 and CG11376, KIAA1058, rat TRG (isoform
2))"
/protein_id="CAC22149.1"
/db_xref="GI:12274836"
/translation="MLVYVQRLNFVNKLASARNITIKIOFMCEDASNAMPVIEGK
SSGPELOEVYAVYHNKSPDEYEVKIKPAKIVVHNLFTYHHSICOOKGASV
ETLLGSMLPTLNERLOTGSCYCLPALEKLPKPNHSMASAEVPLQNPPIKRAEENK
VFNIEVQAVSSVHTODNHLKEFTLCHSLESQVTPPIRVLDOKISEMALEHKLSTII

```

Query Match 3.7%: Score 92: DB 9: Length 181179:

Best Local Similarity 44.3%: Pred. No. 6.3e-05:

Matches 415; Conservative 0; Mismatches 520; Indels 1; Gaps 1;

```
QY 1410 CAACGGCTTCACTTTCAGAGAAAGACCTCGGGACATCGCCGACAGGCTCTGGGC 1469
DB 33943 CACCACTTCCACATCACCACACCTCCACACACACACATCTTCAACATCACCA 34002
QY 1470 AACGCCACATGGGCCACCTCGACAGAGCTGACTACGCCGGAATCGGTCAAGAGCT 1529
DB 34003 CACCTCCACATCACCACACACCTCCACACATCACCACCTCTTCAACATCACCA 34062
QY 1530 CGTCCGCTGGGACACCGCCGACGACAGAGGTGGCATCTGTCGAGCGGGCGC 1589
DB 34063 CACCAACCACTTCCATTCACACACCTCCACATCACCACCTTCCACATCACCA 34122
QY 1590 GACGTGACATGACATGACGCGCCGACATGACGAGGGGGGACGGTGGCTATTTC 1649
DB 34123 CACCTCCACATCACCACATCTTCTTCAACACACACATCTTCAACGTCACACCA 34182
QY 1650 GGCCGAGGCGACAGATCTTGTGACCGGCTGTCGGGGGTGACGCTCGCATTT 1709
DB 34183 CTTCACACACACACCTCTCCACACACACACACCTCTCCACATCACCACTCTTCA 34242
QY 1710 CCAGGGCAGCTTGGCTCTGAGCCTGCGGCGGCGGCTATCGCTCGAGCA 1769
DB 34243 CATCACATCACACCTCCACATCACACACACACCTCCACATCACCACTCTTCA 34302
QY 1770 GAAGACCAACAGCTTCTTACGCGGCTTC-GGATGACCTTTTACGACGAGACA 1828
DB 34303 CACGACACACATCTTACATCACCACACACCTCTTCAACATCACCACTCTTCA 34362
QY 1829 CGGGAGAGCTTCAAGCGGGGCCAAGCTGGAGAGCGAGGAGATCGGGATATCG 1888
DB 34363 CCACCATCACACACCTCTCCACATCACACACCTCTCCACATCACCACTCTTCA 34422
QY 1889 CTGCTACCCGACACAGCGGGGACGCTTATGTCGACCGAGCTGGCATATTCGCT 1948
DB 34423 CCATCACACACACCTCTCCACATCACACACCTCTCCACATCACCACTCTTCA 34482
QY 1949 CCACAGACTCGGGACAGACTTTGGCAAGTCTCCACGCGCGGACCAACCTTACAGA 2008
DB 34483 CCACCATCACACACCTCTCCACATCACACACCTCTCCACATCACCACTCTTCA 34542
QY 2009 TCGCCCTGGGCTGAGGCTGAGACTGAACTGTATGCTTGGACGCGGCGCT 2068
DB 34543 TCACCACTCACACCTCTCCACATCACACACCTCTCCACATCACCACTCTTCA 34602
QY 2069 CAGGGCTGCTCTAGCGCATGAGAGAGGCGGCTCTGAGAGCATTCAGGGCT 2128
DB 34603 CCAGCATCTTCAACATCACACCTCTCCACATCACACCTCTCCACATCACCA 34662
QY 2129 CCCAGGGCTTTCGCTTCATCGACAGCAAGGTGCGGGGACAGCAGCGCGGGC 2188
DB 34663 CCTCCACATCACACCTCTCCACATCACACCTCTCCACATCACCACTCTTCA 34722
QY 2189 AAGTCTAGTGGGACCAAGGCGGGGCTTTTACGCTCAGGGAACGTCGCGGGC 2248
DB 34723 CCTCCACATCACACCTCTCCACATCACACCTCTCCACATCACCACTCTTCA 34782
QY 2249 GCACGGGCGGACTTCTGTCGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2308
DB 34783 CCATCACCAACATTCACACCTCTCCACATCACACCTCTCTCCACATCA 34842
QY 2309 CGAGCAACAGCTGAGGCTGAGGCTTATCCACGA 2344
DB 34843 CCACCTCTCTCACACACCACTTATGCTCCACCA 34878
```

Search completed: May 4, 2003, 11:25:50
Job time : 5257.99 secs

THIS PAGE BLANK (USPTO)

XX Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -

PS Claim 89; Page 3034; 3161pp; English.

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organization of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from *Aspergillus*
 CC *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
 CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
 CC all specifically claimed in the present invention.

XX Sequence 1103 BP; 230 A; 322 C; 307 G; 223 T; 21 other;

Query Match 23.9%; Score 600.4; DB 21; Length 1103;

Best Local Similarity 85.8%; Pred. No. 1e-115;

Matches 810; Conservative 0; Mismatches 96; Indels 38; Gaps 12;

OY 944 GCGGCTTGGCTCGATTTGCAAAAGCAGAAACCTTGTGTTGCTTGTGAACCTTT 1003
 DB 1 GCGGCTTGGCTCGATTTGCAAAAGCAGAAACCTTGTGTTGCTTGTGAACCTTT 60
 OY 1004 GGTGGCCAAATCTCAGCTTTTGGTTCGACGACTCTGGGACACATGAGCCCATCT 1063
 DB 61 GGTGGCCAAATCTCAGCTTTTGGTTCGACGACTCTGGGACACATGAGCCCATCT 120
 OY 1064 GGGCGTGGCGAGCTATCGAGTGAAGACCTATTACTAGACATCTCAACCTCCAAAGCAC 1123
 DB 121 GGGCGTGGCGAGCTATCGAGTGAAGACCTATTACTAGACATCTCAACCTCCAAAGCAC 180
 OY 1124 CGTGGATCAAGAAACACTTATTCGATGTAGAGCGAGTCACCGTCCGATGGTCTCATCA 1183
 DB 181 CGTGGATCAAGAAACACTTATTCGATGTAGAGCGAGTCACCGTCCGATGGTCTCATCA 240
 OY 1184 ACCGCTCGGCTGGATGATGAGTCTCTCGAGATGATGACCAACCGACACCAACTGGC 1243
 DB 241 ACCGCTCGGCTGGATGATGAGTCTCTCGAGATGATGACCAACCGACACCAACTGGC 300
 OY 1244 -TCTAGCGACCGAGTGAATCTTTGGCGGACAGATCTCAACCACTGGGACAGCGGC 1302
 DB 301 TTTTACGCGACCGAGTGAATCTTTGGCGGACAGATCTCAACCACTGGGACAGCGGC 360
 OY 1303 C--ACAATGTGTAATCAATCAAC--TGCGAGAGCGCAT---CGAGAAATTCCTCCGTCA 1355
 DB 361 CCACATATGTGTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 420
 OY 1356 GGACCTGGGCTCTGCACCC--GGCGAAGGAGACTATGGCCGA--GTGAGAGACA--CA 1411
 DB 421 GGACCTGGGCTCTGCACCC--GGCGAAGGAGACTATGGCCGA--GTGAGAGACA--CA 480
 OY 1412 AGCGCTTCACTTT--GCCAGAGAAAGACCTCGGAGATCGCGCAGACAGTCTGGGC 1469
 DB 481 ACGGCTTCACTTT--GCCAGAGAAAGACCTCGGAGATCGCGCAGACAGTCTGGGC 540

OY 1470 AACGCCACATGGGCGCCACCTCGACG--AGCGTGACATACCCCGGAACTCGTCAAGAC- 1527
 DB 541 AACTCCACATGGGCGCCACCTCGACGAGGAGGTGACATACCCCGGAACTCGTCAAGACG- 600
 OY 1528 -GTTCGCGCGTTCGCGACACAC- GCCGGCAGCGACACAGGTGGCCATCTGTCGACGCGC 1585
 DB 601 GTTCGCGCGTTCGCGCGACACACAGCGCGCGACGCAACAGGTGGCCATCTTTTCCGACG 660
 OY 1586 GCGCGCGTGGAGCATGCACTACGCGCGCGCGACACGTCAT----- 1625
 DB 661 GCGCGCGCGTGGAGCATGCACTACGCGCGCGCGCGACACGTCAT----- 720
 OY 1626 GAACGGCGGCGAGCGTCCCTATTCGCGCGCGCGCGACAGATCTGTCGACGCGCTC 1685
 DB 721 AAACGGCGGCGGCGTCCCTATTCGCGCGCGCGCGACAGATCTGTCGACGCGCTC 780
 OY 1686 GTTCGCGCGTTCGCGCGTTCGCGACAGTTCGAGGCGACGTTGCTCGTTCGAGCCTGCGCC 1745
 DB 781 GTTCGCGCGTTCGCGCGTTCGCGACAGTTCGAGGCGACGTTGCTCGTTCGAGCCTGCGCC 840
 OY 1746 GGGCGCGTTCGCGCGTTCGCGACAGTTCGAGGCGACGTTGCTCGTTCGAGCCTGCGGATC 1805
 DB 841 GGGCGCGTTCGCGCGTTCGCGACAGTTCGAGGCGACGTTGCTCGTTCGAGCCTGCGGATC 900
 OY 1806 GACCTTTTACGTACGACAGACACCGCGACAGCTTACGCGCG 1849
 DB 901 GACCTTTTACGTACGACAGACACCGCGACAGCTTACGCGCG 944

RESULT 2

AAF15118
 ID AAF15118 standard; cDNA; 406 BP.

XX AAF15118;

XX 13-MAR-2001 (first entry)

XX Trichoderma reesei EST SEQ ID NO:7641.

XX Multiple gene expression: filamentous fungal cell; EST:

XX expressed sequence tag; *Fusarium venenatum*; *Aspergillus niger*;

XX *Aspergillus oryzae*; *Trichoderma reesei*; identification; recombination;

XX culture condition; environmental stress; spore morphogenesis;

XX metabolic pathway engineering; catabolic pathway engineering; ss.

XX Trichoderma reesei.

XX WO200056762-A2.

XX 28-SEP-2000.

XX 22-MAR-2000; 2000MO-US07781.

XX 22-MAR-1999; 990US-0273623.

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX (NOVO) NOVO NORDISK AS.

XX Berta RM, Rey MM, Shuster JR, Kauppinen S, Clausen JG, Olsen PB;

XX WPI; 2000-594572/56.

XX Monitoring differential expression of genes in filamentous fungal cells

XX uses fluorescence-labeled nucleic acids isolated from the cells and a

XX substrate of expressed sequence tags -

XX Claim 89; Page 3084-3085; 3161pp; English.

XX The present invention describes a method for monitoring differential

XX expression of genes in a first filamentous fungal (FF) cell relative to

XX expression of the same genes in one or more second filamentous fungal

XX cells. The method uses fluorescence-labeled nucleic acids isolated from

CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic, or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF1247 represents ESTs from
CC *Fusarium venenatum*, AAF1248 to AAF1853 represents ESTs from *Aspergillus*
CC *niger*, AAF1854 to AAF1878 represents ESTs from *Aspergillus oryzae*, and
CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
CC all specifically claimed in the present invention.

SO Sequence 406 BP; 75 A; 141 C; 123 G; 64 T; 3 other;

Query Match 13.0%; Score 326.6; DB 21; Length 406;
Best Local Similarity 93.6%; Pred. No. 8.9e-59;
Matches 382; Conservative 0; Mismatches 22; Indels 4; Gaps 4;

OY 1871 GGACGATCCGGATATCCGTGCTACCCACGACCGCGGGGACGCTGTATGCTGCACCG 1930
DB 1 GGACGATCCGGATATCCGTGCTACCCACGACCGCGGGGACGCTGTATGCTGCACCG 60
OY 1931 AGCTGGCATATTCGCTCCACAGACTCGGGACGACTTTGGGCAAGTCTCACCGGCC 1990
DB 61 AGCTGGCATATTCGCTCCACAGACTCGGGACGACTTTGGGCAAGTCTCACCGGCC 120
OY 1991 TCACCAACACCTACCAAGATGCGCTGCTGGGCTGAGGCTGCAAGTCTGAACTGTATG 2050
DB 121 TCACCAACACCTACCAAGATGCGCTGCTGGGCTGAGGCTGCAAGTCTGAACTGTATG 180
OY 2051 CCTTGGGACCGCGCGTGGGAGGCGCTGCTAGCGCCAGTGGAGAC-AGCGGCGGCTCC 2109
DB 181 CCTTGGGACCGCGCGTGGGAGGCGCTGCTAGCGCCAGTGGAGACAGAGCGGCGCTCC 240
OY 2110 TGGACGAGACATCC-AGGCTCTCCAGGGCTTCGCTCCATGACAGACCAAGGTCGCCGG 2168
DB 241 TGGACGAGACATCCAGGGCTTCGCTCCAGGGCTTCGCTCCATGACAGAC-ACCAAGGTCGCCGG 299
OY 2169 CAGCGGACGACCGCGCGGCAAGTCTACGTGGGACCAAGCGCGCGGCTCTTTTACGC 2228
DB 300 CAGCGGAAAG-ACCGCGCGGCAAGTCTACGTGGGACCAAGCGCGGCGGCTTTTACGCT 358
OY 2229 TCAGGGAACCGTGGGCGGCGGACGCGGGAATCTCTGCTGACCA 2276
DB 359 TAAGGAACCGGCGGCGGCGGACAGCGCGGAATCTCTGCTGACCA 406

RESULT 3
AA199683/C

ID AA199683 standard; DNA: 4403765 BP.

XX AA199683;

DT 15-JAN-2002 (first entry)

DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

XX Mycobacterium tuberculosis; strain H37Rv; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.

OS Mycobacterium tuberculosis.

PN US6294328-B1.

XX

PD 25-SEP-2001.

XX 24-JUN-1998; 98US-0103840.

XX 24-JUN-1998; 98US-0103840.

PA (GENO-) INST GENOMIC RES.

PI Fleischmann RD, White OR, Fraser CM, Venter JC;

DR WPI; 2001-647261/74.

PT Evaluating strain variation of *Mycobacterium tuberculosis*, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where *M. tuberculosis* strains CDC
PT 1551 and H37Rv differ.

PS Claim 4: SEQ ID NO 2; 3bp + Sequence Listing: English.

CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen.
CC *Mycobacterium tuberculosis* or related *Mycobacterium* by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of *M. tuberculosis* strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC *M. tuberculosis* and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPRO
CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.

SO Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 3.5%; Score 88; DB 22; Length 4403765;
Best Local Similarity 43.0%; Pred. No. 2e-08;
Matches 433; Conservative 0; Mismatches 575; Indels 0; Gaps 0;

OY 1391 TGGCGGCACTGGAGAGGACGACGACGCTTCCATCCGACGAGAAAGACGCTGGGACAT 1450
DB 3937116 TGGCGGCGGCTGGCGCGCGCTGCGCTCCACCTGCGCGCGCGCGCGCGCGCGCG 3937057
OY 1451 CGCGGAGAGGCTGCGGAGACGCGCCACATGGGCGACCTGACGAGCTGACGACGCGG 1510
DB 3937056 CGAGGTTACCGGCTTGTTCACCGGAGCGCGGAGCGGAGCGGACCGTCCAGCGGTCGGG 3936997
OY 1511 GGAACCTGGTCAAGAGAGCTGCGGCTGCGGACACCGCGGCGGACGACGAGTGGCCA 1570
DB 3936996 TCGCGGCGGAGGCGCGCGGACCGCGGCTGCGCGGCGGCGGCGGCGGCGGCGG 3936937
OY 1571 TCTCGTCCAGCGCGCGGCGGAGTGGAGTACGACGACGCGGCGGCGGCGGCGGCGG 1630
DB 3936936 CCACCGGTCGCGCGCTTGGCGCGCGCGCGCGCGCGCTTGGCGGCGGCGCGCGGCTTACCGGCGG 3936877
OY 1631 GCGGACGAGTGGCGCTATTCGCGGCGGAGCGGACGACGATCTGATGGAGCGCGCTGCGCG 1690
DB 3936876 GCGCGCGGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3936817
OY 1691 GCGTGCAGCGCTGCGGAGGTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1750
DB 3936816 GCACCGCGGCGGACGACGCTTGGACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3936757
OY 1751 CGGTCACTGCGCTGCGGAGGAGGACGACGAGGCTTGTTCAGCGCGGCGGCTCGGATGACCT 1810
DB 3936756 CGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3936697
OY 1811 TTACGTCAGAGGACGACGCGGAGCGGCTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1870
DB 3936696 CTGCGCGGCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3936637
OY 1871 GGACGATCCGGATATCCGTGCTACCGACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1930

Db	20073	CCGAGCGACACCGGGCCCGCCCGGCAACCGTTCACACGGGACTCCCTCTTCCACGTGAGTGGAA	201322
QY	1451	CGCCGACAGAGGCTGTGGGCAACGCCACATAGGGCCACCTCGACAG6CGTGCATAGCGCG	1510
Db	20133	CCCCCTCCAGGGCGCGCCCGGGCCCGGACCGGACCGCAGCGTGCCTGCTCGGCCCCGAAAC	201922
QY	1511	GGAACTCGGTCAAGAGCGCTGCTGCCCTGTCGGCAACACGGCCGGCAACGCAAC6AGTGGCCA	1570
Db	20193	CGAGCGCCCTCGCCGACACCCCTCCCGGCACACGGGATCCGACACCGCCGCCCGCCCGAAC	202522
QY	1571	TCTCGTCCGACGGCGGCGGCGAGAGTGGAGCATCGATACGGCGGCCCAACAGTCCATGAAGC	1630
Db	20253	TGGCGGCGCTCGCCGACCGCCGAAGGGGCCCGTCCCGACTGGTGGTCAACACCTCACCBA	203122
QY	1631	GGCGACGGTGGCCCTATTTCGGCCGACGGCGACACAGCATCCTCTGATCGAACCGCCTGTCCG	1690
Db	20313	CCACCCCGGGGGCCCCCGCTCCCGGCGCGGCGACGGCCACGACCGCCGCGCGCTCTGCGCC	203722
QY	1651	GGCTGCACCGCTCGCAGATTCCAGGGCAGCTTTGCCCTCCGTCTCGAGCCTGCCCGGGGCG	1750
Db	20373	TGCGCCCAACAGTGGGTGCGCCGACGACCGCTT---GGCGAGCCCGCGCTGGTCTCTGTCA	204292
QY	1751	CCGTATGATGCGCTCGGACAAGAACCAAGCATAGGGTCTTCAACCGCGGCTCCGGATGAGACT	1810
Db	20430	CCCGGGGGGCGACCGACCGGACCCCGCCGCGCGCGCGCGCGCGCGCGCGCTGATTCGGA	204892
QY	1811	TTTACGTAGCAGCAAGGACACCGGACAGCTTCACGCGCGGGCCCAAGCTGGGACAGCGAG	1870
Db	20490	CCGGCCGACCGAGAACCCCGGCGCTTTCGCGCTCTGAGACTCGCCCCGACACCGGGC	205492
QY	1871	GGAGCATCCGGGATATTCGCTGCTCAACCCGACCGACCGCGGGGACAGTTGATGTCTGACCG	1930
Db	20550	GGCCCGACCCCGAGACCCCTGGGCAACCGGCTGGGCGGCAAGCAGCAGCAGCGCCGACCTCG	206092
QY	1931	ACGTGCGCATATTTCGCTTCACAGACTCGGGGACAGCACTTTGGGCAAGTCTCAACCGGCC	1990
Db	20610	CCGTGCGGGGACCGACGACGTGCAGCGCGCGCGCGCTGCTGCTCCCGCTCGGCAACCGAAC	206652
QY	1991	TGACCACACCTACCGAGATCGGCTTGGGTGTGGGCTCAGGCTCGAATGGAACTGTATG	2050
Db	20670	CCACCACTTGGAAACCGGACGGCACCGTCTGTATCAGCGGCGGCAACGGCGGCTGGGGG	207292
QY	2051	CGTTGGGACCGGGCCCGGTACAGGGGGTCCGCTTACGGCAGTGGAAACAGGGCGGCTCT	2110
Db	20730	CGGTCTCGCCCGCCACCTGTGTGCGCACCAAGGGGTCCGCACTGCTGTCTGCGCAGGC	207892
QY	2111	GGACGAGCATCCAGGGGCTCCGAGGGCTTCGGG---TGCATCGACAGCAACAGTGTGCGCG	2167
Db	20790	GCGCGGGCCCGGCGCCGACGGCGCGGACGACTACAGGGCCGCAACTCACCGGGCTTCGGCG	208492
QY	2168	GCAGCGGACGACCGCGGGGCAAGTCTAAGTGGGACCAACAGCGCGCGGGCGTCTTTTAA	2227
Db	20850	CCACGCTTCACATCGCGCGCTCGAGACGTGCGGAGACCGGCGCGCCCTCGCGGACGTGCTCG	209092
QY	2228	CTCAGGGAACGCTGGG 2243	
Db	20910	GCACCGTCCCGGCGG 20925	

RESULT 5
AAD17186
ID AAD17186 standard; DNA: 125401 BP.
XX
XX
AC AAD17186;
XX
DT 29-NOV-2001 (first entry)
XX
XX
DE Streptomyces noursei nystatin PKS gene cluster DNA.
XX
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; ds.
XX

OS	Streptomyces noursei.
XX	Location/Qualifiers
XX	Key
XX	6337..34771
XX	/tag= a
XX	/product= "NysI complete protein"
XX	34792..51099
XX	/tag= b
XX	/product= "NysJ protein"
XX	51155..57355
XX	/tag= c
XX	/product= "NysK protein"
XX	57503..58687
XX	/tag= d
XX	/product= "NysL protein"
XX	complement (58786..58980)
XX	/tag= e
XX	/product= "NysM protein"
XX	/note= "CDS does not include start codon"
XX	complement (59045..60241)
XX	/tag= f
XX	/product= "NysN protein"
XX	/note= "CDS does not include start codon"
XX	complement (60238..61296)
XX	/tag= g
XX	/product= "NysD2 complete protein"
XX	120628..121308
XX	/tag= h
XX	/product= "NysR4 (long) protein"
XX	WO200159126-A2.
XX	16-AUG-2001.
XX	08-FEB-2001; 2001WO-GB00509.
XX	08-FEB-2000; 2000GB-0002840.
XX	10-APR-2000; 2000GB-0008786.
XX	14-APR-2000; 2000GB-0009387.
XX	(UNNO-) UNIV NORGES TEKNISK NATURVITSENSKAPELIGE.
XX	(SMFE) SINTEF SPITFELSEN IND TEK FORSK.
XX	(ALPH-) ALPHARMA AS.
XX	(SINV-) SINVENT AS.
XX	(DZIE/) DZIEGLEWSKA H.
XX	(ZOTC/) ZOTCHEV S B.
XX	(SEKU/) SEKUROVA O N.
XX	(FJAE/) FJAEVRIK E.
XX	(BRAU/) BRAUTASET T.
XX	(STRO/) STROM A R.
XX	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX	Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX	WPI: 2001-557614/62.
XX	P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX	AAE10149, AAE10150.
XX	New nystatin polyketide synthase polynucleotides and polypeptides,
XX	useful as antibiotics and antifungals -
XX	Claim 1; Page 188-254; 266pp; English.
XX	The present invention relates to the cloning and sequencing of the gene
XX	cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX	involved in the biosynthesis of the macrolide antibiotic nystatin.
XX	The nystatin PKS is useful as antifungal antibiotics. The present
XX	sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX	Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
XX	Query Match 3.5%; Score 87.2; DB 22; Length 125401;
XX	Best Local Similarity 45.0%; Pred. No. 1.6e-08;

Matches	412;	Conservative	0;	Mismatches	498;	Indels	6;	Gaps	2
Oy	1331	ACGGCATCGAGAAATTTCCTTCACAGAGACCTTGCTTCACCCGGCGAAGCGACTAT							1390
Db	80274	ACGAGCGCGCCCGCGGACCTGCTGTCTCGTAGACTCCCTGGTCTCTCCCGAGGTCCCG							80333
Oy	1391	TGGCGCGAGTGGAGACGACAAACGGGCTTACCTTTGGTCCAGCAAGAAAGACCTCGGGACAT							1450
Db	80334	CCGAGCGACCCCGCGCCCGCGGACCGTTCACCGGACTTCCTTTCCAGTGTGATGGA							80393
Oy	1451	CGCCGACAGCGGTCTGGGCAACGCCCATATGGGCCACTTCGACGAGCGTGCATACGCCG							1510
Db	80394	CCCCGCTCCAGGGCGCCCGCGGCGCCGACCGGACCGTGCCTCCGCTCTCGGCCCGGACC							80453
Oy	1511	GGAATCTCGGTCAAGAACGCTGTCCCTGTGGCAACACCCCGGCAAGCAACAGTGGCCA							1570
Db	80454	CGGAGCGCTTCCCGACACCTCTCTCGGGCCACCGGCTCCGGACACGCGCCCGCCCGGACC							80513
Oy	1571	TCTCTCGACGCGCGCGCGGACGAGTGGAGCATGAGTACGCGCGCCGACAGTCCATGAAG							1630
Db	80514	TGGCGCGCTCGCCCGACGCGCGAAGGGCCCGCTCCCGACCTGGTGTACACACCTTCACCA							80573
Oy	1631	CGCGACAGGTGGCTTATTTGGCGCGACGCGGACACGATCCTTGTGTGACCCCTCTCGCG							1690
Db	80574	CCACCCCGCGCGCCCGCCCGTCCCGGACGCGCGCGACCGCACACCGCGCGCGCTCTCGGCC							80633
Oy	1691	GCGTCCAGCGCTTCGCGAGTTTCCAGGGCAGCTTTTGCCCTCGCTTCGAGACTGCGCGCGGGCG							1750
Db	80634	TGCGCCCAACAGTGTGCTCGCGCGACGACCGCTT---CGCCACGCGCGCTGTGTCTGTGTA							80690
Oy	1751	CCGTATCGCTTCGACAGAAGAACCAACAGCGCTTTTACGCGCGGCTCCGCGATCGACT							1810
Db	80691	CCCGGCGCGCCACCGACGCGACCGCCCGCGCGCGCGCGCGCGCGCGCGCGCTGTATCGCA							80750
Oy	1811	TTTACGTACGACGAAGACACCGGACAGCTTACAGCGCGGGCGCCCAAGTGGGACGGCAG							1870
Db	80751	CCGCGCGCACGAGAACCCCGCGCGCTTTCGCTCTCTGACCTCGCCCGACACCGGCGC							80810
Oy	1871	GGACGATCCGGGATATCGCTGCTCAACCCGACCGCGCGGACGTTGTATGTCTGACCG							1930
Db	80811	GGCGCGACCCCGAAGACCTTGGCGACCGCCCTGGCGCGCCAGCCAGCAGAGCGGACCTCG							80870
Oy	1931	ACGTGGGCTATTTCCGCTCCACAGACTCGGGGACAGACTTTTGGCCAAAGTCTCACCGCCC							1990
Db	80871	CGGTCCGGCGCACGACGTCAGCGCGCGCCCGCTGGCTGTCCCGCTCGCCACCGAAGC							80930
Oy	1991	TGACCAACACTACAGATCGCTTGAGTGTGGGCTCAGGCTCGAAGTGAAGCTTATG							2050
Db	80931	CCACGACCTGGAACCCCGGAGCGGACACGCTGTATCAGCGCGCGGACACGGCGGCTGGGG							80990
Oy	2051	CCTTGGGACACCGGCCGCTCAGGGGCTCGCCTTACGCCAGTGGAGACACGCGCGCTCCT							2110
Db	80991	CGGTCTCGCCCGCGCACCTGTGTGCGCACACGAGGCGTCCGCCACGCTGTCTGCGCAGCG							81050
Oy	2111	GGACGACATCTCAAGGAGCTCCCAAGGGCTTGGGCTTCCATGTGACAGACCAAGTGTGGCG							2167
Db	81051	GCGCGGGCGCGCGCGCGCGCGGACGCGCGCGACGACTGACGGCCGCAACTACCGGGCTGGGG							81110
Oy	2168	GCAGCGGACGACCGCGCGGGCAAGTCTACGTGGGACCAAGAGCGCGGGCGCTTTTACG							2227
Db	81111	CCACGCTCCACTTCGCGCGCTCGCTGACGCTGCGGACCGCGCGCGCTCGCGCGACTGCTCG							81170
Oy	2228	CTCAGGGAACCGTCCG 2243							
Db	81171	GCACGCTCCCGCGCGG 81186							

RESULT 6

AA064206

ID AA064206 standard: cDNA, 1833 BP.

XX AA064206;

XX

DT	18-NOV-1994	(first entry)
DE	snaB gene encoding enzyme in streptogramin biosynthetic pathway.	
XX		
KW	Antibiotic; streptogramin; snaA; snab; snaC; biosynthesis; enzyme;	
KW	biosynthetic pathway; Streptomyces pristinaespiralis; snbA; snbR;	
KW	ds.	
XX		
OS	Streptomyces pristinaespiralis.	
XX		
FH	Key Location/Qualifiers	
FT	CDS 103..1689	
FT	/+tag= a	
PN	FR2696189-A.	
PD	01-APR-1994.	
PF	25-SEP-1992; 92FR-0011441.	
PR	25-SEP-1992; 92FR-0011441.	
PA	(RHON) RHONE POULENC RORER SA.	
PI	Blanc V, Blanchet F, Crouzet J, Jacques N, Lacroix P;	
PI	Thibaut D, Zagorec M;	
DR	WP1: 1994-128286/16.	
DR	P-PSDB; AAR54206.	
XX		
PT	DNA involved in streptogramin antibiotic biosynthesis - for	
PT	prodn. or bio-conversion of streptogramin(s) or prodn. of	
PT	streptogramin intermediates, derivs. or hybrid antibiotics	
XX		
PS	Claim 2: Page 62-64; 83pp; French.	
XX		
CC	The snaB gene product is involved in the biosynthesis of	
CC	streptogramins, antibiotics active against Gram-positive bacteria.	
CC	The identification of the sequences encoding the enzymes involved	
CC	in the biosynthetic pathway means that they can be isolated and	
CC	manipulated. Mutant microorganisms in which a step in the	
CC	streptogramin biosynthetic pathway is blocked can be cultured to	
CC	produce streptogramin intermediates, which may later be converted	
CC	to streptogramin derivatives. Recombinant cells may also be used	
CC	for the bioconversion of streptogramins from one form to another or	
CC	for the production of hybrid antibiotics.	
XX		
SQ	Sequence 1833 BP; 232 A; 802 C; 564 G; 235 T; 0 other;	
Query Match	3.1%; Score 78.4; DB 15; Length 1833;	
Best Local Similarity	46.1%; Pred No. 5.6e-07;	
Matches 339; Conservative	0; Mismatches 391; Indels 6; Gaps 2	
OY	1483 GCCACCTCGAGCAGCGGTGACTACGACCAGCACTGGTCAGAGCGTCTCGCGTGCGC 1542 	
Db	820 GCCATCATCTCGGCCGCCGACACAGCGTGCAGCGCCCCGACAGTCTCTCGCGCGCC 879 	
OY	1543 AACACCGCCGCGACGACMACAGSTGGCCATCTGCTCGACGGGGCGCACGTGGACATC 1602 	
Db	880 GTCGGGCGCGCGCGCTCATCCCGCGTTGTCCTGTGGAACTGCACACCCCACCCCATG 939 	
OY	1603 GACTACGCGGCGCGACGACGTCATGAAGCGCGGACAGGTGGCTATTGCGCGACGGCGAC 1662 	
Db	940 CTCGACATGAGGCTTTTACACCGACGGCGCTTTCACAGGGCCGCTCGCGCGGAGTG---C 996 	
OY	1663 ACGATCTCTGTGTCGACCGCTCTGTCGGCGTGCACAGCGTGCAGTTCAGGGCAGACTTT 1722 	
Db	997 TCGGTCTTCGCGCATGCGCGGGGTGCGCTTCTCGTCACACGACGACACTTCAACCTCGTCTC 1056 	
OY	1723 GCCTCCGTCGACGCTCGCGCGGGGCGCGGTTCATCTGCTCGACACAAGAAGACCAACAGC 1782 	
Db	1057 GGCTTACGACGCCCTTCAGGCCCTGCGCCACCGGCCCACTGGCTTTGACATCGTCGCC 1116 	

DE Human METH1 related EST Z46970.

KM Human, MEHL: METH: anti-angiogenic; metalloprotease thrombospondin;
KM cancer: diagnosis; hyperproliferative disorder; autoimmune disease;
KM angiogenesis inhibitor; abnormal wound healing; inflammation;
KM rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
KM diabetic retinopathy; macula degeneration; haemangioma; detection;
KM arterial-venous malformation; immune deficiency; ss.

OS Homo sapiens.

PN W09937660-A1.

PD 29-JUL-1999.

PF 22-JAN-1999; 99WO-US01313.

PR 23-JAN-1998; 98US-0072298.

XX

PA (HAST/) HASTINGS G A.

XX

XX

XX

PT treating hyperproliferative

XX
XX

XX

CC AAZ32000 and AAZ32001 encode, and AAAY49501 and AAAY49502 represent, human
CC metalloproteinase thrombospondin (METH1) proteins METH1 and METH2
CC respectively. METH1 and METH2 have been found to be potent inhibitors of
CC angiogenesis both in vitro and in vivo. They can be used for treating
CC cancer and other disorders related to angiogenesis including abnormal
CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
CC endometrial bleeding disorders, diabetic retinopathy, some forms of
CC macula degeneration, haemangiomas, and arterial-venous malformations.
CC They may be useful in treating deficiencies or disorders of the immune
CC system, by activating or inhibiting the proliferation, differentiation,
CC or mobilisation (chemotaxis) of immune cells. The etiology of these
CC immune deficiencies or disorders may be genetic, somatic, such as
CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
CC toxins), or infectious. They can also be used to treat inflammatory
CC conditions, both chronic and acute conditions. The products can also be
CC used for detection and diagnosis. AAZ32002 to AAZ32080, and AAAY49503 to
CC AAAY49511 represent sequences given in the exemplification of the present
CC invention.

SQ Sequence 3015 BP; 674 A; 1079 C; 831 G; 431 T; 0 other;

Query Match 2.88; Score 69.8; DB 20; Length 3015;

Matches	396;	Conservative	0;	Mismatches	517;	Indels	6;	Gaps	1;
---------	------	--------------	----	------------	------	--------	----	------	----

0Y	1405	GAGCAAGGGTTCACTTTGGCAGCAGAAAGACACTTGGGACATCGCGCAGACGGTC	1464
Db	1478	GCCACCGACGACGACGATTACCAACACAGAGAGCGAAGGTTACCGCCACCGACGACGC	1537
0Y	1465	TGGGCAACGGCCACATGTGGCCACCTCTGACGAGCGTGTGACTACCGCGGAACCTGGTCAAG	1524
Db	1538	AGCGACCGCTACCGACGACGACGAGGGTACCGCCACACGACGACGAGACGCTACACAC	1597
0Y	1525	AGCGTGTCCGGCGTGGCCACACACGGCGCGGACGACACAGTGGCCATCTCTGTCCAGCGC	1584
Db	1558	AGCAGCAGCAGCAACGCTTACCAACACACGACGACGCGAAGGGTTACCAACCGACGACGACGC	1657
0Y	1585	GGCGCGACGTGGAGCATGCATACGCGGGCGACGACGCTTCATGAACGGCGGCAACGGTGGCC	1644

Db	1658	GAGCCTACACACACACACACACACACGCTACCACCACACAGTACAGCAGGAGGATACCGCCACC	1717
Qy	1645	TATTTCGGCCCGACGGCCGACAGGATCCTTGTGTGACCGCCCTCTCGCGGTGACAGGCTCG	1704
Db	1718	AGCAGACAGCCGCTATCCACACACACAGAGGAGGGATACCGCCACACAGCAGCAGCAGCTT	1777
Qy	1705	CAGTTCACAGGACAGCTTTGCCCTCCGCTGTGAGCTCTGCCCGGGGGCCGTATGGCTCTCG	1764
Db	1778	ACCAACACACACACACGAGGGTATCCCGCACACAGCAGCAGGACGCTACCAACACACAGCAGC	1837
Qy	1765	GACAAAGAACCAACACAGCTCTTTCACGGCCGGCTCCGATTCGACCTTTTACGTCAGCAAG	1824
Db	1838	AGCAGAGGTATCACACACACACACACAGACGAGCTACCAACAGCAGACGAGAGGTATCGCC	1887
Qy	1825	GACACCGGACAGCAGCTTTCACGCGCCGGGCCCAAGCTTGCGACGCGCAGGAGCAGTCCGGAT	1884
Db	1898	ACCAACACAGCAGCGACGGCTATCCACACAGAGAGAGGAGGATACCAACAGCAGCAGCAGC	1957
Qy	1885	ATTCGCTGCTACCCGACACCCGGGGCAGCTGTGTATGTCTGCACGACGCTGGCATATTTC	1944
Db	1958	GACCTATCACACACACACACAGCAGCAGCTTACCACACACAGCAGCAGCAGGAGGTATCCGCC	2017
Qy	1945	CGCTCCACAGACTCTGGGACAGACCTTTGGCCAAAGTCTCCACGGCCCTATCAACAACACTAC	2004
Db	2018	ACCAAGCAGCAGCAGCAGCTTACACACAGCAGCAGCAGGATACCGCCACCAACAGCAGCAGC	2077
Qy	2005	CAGATTCGCCCTGGGTGTGTGGGCTCAGGCTCAGACTGMACTCTATATGCCCTTGGCACCGC	2064
Db	2078	GCTTACACACACACACACACACAGGATACCAACAGCAGCAGCAGGAGCTTACCAACCAACAGC	2137
Qy	2065	CCGTACAGGGGCTCGCTCTTACGCCCAAGTGGAGA----CAGCGGGCGCTCTGTGACGGAC	2118
Db	2138	AGCAGCAGGGGTATCCGCACACACACAGAGGAGCGTATCAACACAGCAGCAGCAGCAGGAGGT	2197
Qy	2119	ATTCAGGGCTCCCAAGGGCTTCGGCTCCATCTGACAGACCAAGTGGCGGACAGCGCACAGC	2178
Db	2198	ACCAACACACACACAGAACTGTACCTACCAACAGCAGAGGAGGATACCGCCACCAACCAACG	2257
Qy	2179	ACCGCGGGGCAAGTCTACGTGAGGACCAACAGCGCGGGGGGCTCTTTTACGCTCAGGGAAC	2238
Db	2258	AGCGACGCTACCAACACACAGCAGCAGCAGCAGGAGGTACCAACACAGCAGCAGCAGCAGCTTAC	2317
Qy	2239	GTGCGCGCGGACAGGCGGGGAACTCTCTGTGACCAAGCAGACGACAGCAGTACCTCT	2298
Db	2318	ACCAAGCAGCAGCAGGAGGTATCCCGCACACACAGCAGCAGCAGCTTACCAACAGCAGCAGC	2377
Qy	2299	TTCGCGCACTGTGACACCA 2317	
Db	2378	GAGGGTACCAACACACAGCA 2396	

RESULT 10

ID AAC90084 standard; DNA; 3015 BP.

AC AAC90084;

DT 19-MAR-2001 (first entry)

DE 246970 cDNA clone

KM METH; metalloproteinase; thrombospondin; angiogenesis inhibition
KM cancer therapy; benign tumour; ocular angiogenic disease;
KM rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KM vasculogenesis; granuloma; hypertrophic scar; nonunion first
KM scroderoma; trachea; vascular adhesion; myocardial angiogenesis
KM coronary collateral; cerebral collateral; arteriovenous malformation
KM ischaemic limb angiogenesis; Osler-Weber syndrome; wound granulation
KM plaque neovascularisation; telangiectasia; haemophilic joint;
KM angiodioma; fibromuscular dysplasia; expressed sequence tag;
KM Cohn's disease; atherosclerosis; birth control; ss

OS Unidentified.

XX WO200071577-A1.
 PN 30-NOV-2000.
 PD 25-MAY-2000; 2000WO-US14462.
 XX 25-MAY-1999; 99US-0318208.
 XX 20-JUL-1999; 99US-0144882.
 PR 10-AUG-1999; 99US-0147823.
 PR 13-AUG-1999; 99US-0373658.
 PR 22-DEC-1999; 99US-0171503.
 PR 22-FEB-2000; 2000US-0183792.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (IREU/) IREUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 PA (JONA/) JONAK Z L.
 PA (TRUL/) TRULLI S H.
 PA (FORN/) FORMALD J A.
 PA (TERR/) TERRETT J A.
 XX Irueia-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 PI Formwald JA, Terrett JA;
 XX WPI: 2001-025136/03.
 DR METH1 and METH2 polynucleotides and encoded polypeptides, used to
 PT inhibit angiogenesis in the treatment of disorders such as cancer,
 PT rheumatoid arthritis and psoriasis.
 PT
 PS Claim 7: Pages 661-663; 768pp; English.
 XX
 CC The present invention relates to human METH1 and METH2, (ME for
 CC metalloproteinase and TH for thrombospondin; see AAB50002 and AAB50003).
 CC The present sequence is an expressed sequence tag (EST) for METH. METH
 CC can be used for inhibiting angiogenesis in an individual, and for
 CC treating cancer, benign tumours, an ocular angiogenic disease,
 CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
 CC vasculogenesis, granulomas, hypertrophic scars, nonunion fractures,
 CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
 CC coronary collateralis, cerebral collateralis, arteriovenous malformations,
 CC ischaemic limb angiogenesis, Osler-Weber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiodiroma,
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH can also be used in birth control. METH can also
 CC be used in diagnostic methods for the prognosis of cancer.
 CC
 XX
 SQ Sequence 3015 BP: 674 A: 1079 C: 831 G: 431 T: 0 other;
 Query Match 2.8%; Score 69.8; DB 22: Length 3015;
 Best Local Similarity 43.1%; Pred. No. 3.7e-05;
 Matches 396: Conservative 0; Mismatches 517; Indels 6; Gaps 1;
 QY 1405 GAGGACAGGCTTACCTTCCAGCAGAAAGACCTTCCGACATCCCGCAGAGCTC 1464
 DB 1478 GCGACACAGCAGCAGCAGCTTACACACACAGCAGCAGGAGTACCCCGACAGCAGC 1537
 QY 1465 TGGGCAAGCCCAATGGGCACTTCGACGAGCGTTCGACCGCGGAATCGGTCAAG 1524
 DB 1538 AGCAGACGTAACACAGCAGCAGCGAGGGTACCGCAGCAGCAGCAGCAGCTACACAC 1597
 QY 1525 ACCGTCGTCCGGTGGGCAACACCGCGCAGCAGCAAGAGTCTCTCTCCAGCGC 1584
 DB 1598 AGCAGCAGCAGCAGCAGCTTACACACACAGCAGCAGGAGGTACCAACGACAGCAGC 1657
 QY 1585 GCGGAGCAGCTGAGCATCAGCGCGGCGACAGCTTCATGAAGCGCGCAGCGGTGGCC 1644
 DB 1658 GACGCTACCAACAGCAGCAGCAGCTTACACACAGCAGCAGGAGGTACCGCCACCC 1717

QY 1645 TATTCGCGCAGGCGGACAGCATTCCTTGTCGACACCGCTCTGCGCGGTGAGCGCTCG 1704
 DB 1718 AGCAGCAGGAGCTTACACACACAGCAGCAGGAGGTACCGCGCAGCAGCAGCAGCTT 1777
 QY 1705 CAGTTCAGAGGAGCTTTGCTCCGTCTCGAGCTTGCCCGCGGCGCTCATCGCTTCG 1764
 DB 1778 ACCACACAGCAGCAGGAGGTACCGCCACACAGCAGCAGCAGCAGCAGCAGCAGCAGC 1837
 QY 1765 GACAGAGACACAGCAGGCTTCTACGCGCGCTCCGAGTTCAGCTTTTACGTGCAAG 1824
 DB 1838 AGCAGAGGTACACACAGCAGCAGCAGCAGCAGCTTACACACAGCAGCAGGAGGTACCGCC 1897
 QY 1825 GACACCGGAGCAGCTTCAACGCGCGGCCCAAGCTGGCGACGCGAGCATTCGGGAT 1884
 DB 1898 ACCACAGCAGCAGCAGCTTACACACAGCAGCAGCAGCAGGAGGTACACAGCAGCAGCAGC 1957
 QY 1885 ATCGGTGTCTACCCACACACCGCGGCGCAGCTTGATGTCTGACGAGCTGGCATATTC 1944
 DB 1958 GACGCTACACACAGCAGCAGCAGCTTACACACACAGCAGCAGCAGCAGGAGGTACCGCC 2017
 QY 1945 CGCTCCACAGACTTCGCGCAGCAGCTTTGGCCAAAGTCTCCACCGCCTTGACCAACACTAC 2004
 DB 2018 ACCACAGCAGCAGCAGCTTACACACAGCAGCAGCAGGAGGTACCGCCACACAGCAGCAGC 2077
 QY 2005 CAGATCGCCCTTGCGGTGGGCTCAGGCTTCGAACCTGTAATGCTTGGCAGCGGC 2064
 DB 2078 GCTACACACAGCAGCAGCAGCAGGAGGTACACACAGCAGCAGCAGCAGCAGCTACACAGC 2137
 QY 2065 CCGTCAAGGCGCTCGGCTCTACCGCCAGTGAAG------CAGGGGGCGCTCTGGAGCGAC 2118
 DB 2138 ACCACGAGGAGTACCGCCACACAGCAGCAGCAGCTTACACAGCAGCAGCAGCAGGAGGT 2197
 QY 2119 ATCCAGGAGCTCCAGAGGCTTGGCTTCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2178
 DB 2198 ACCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2257
 QY 2179 ACCGCGCGGCAAGTCTACGTGGGCAACAGCGCCGCGGCGCTTTTACGCTCAGGGAAC 2238
 DB 2258 ACCGAGCGTACACACAGCAGCAGCAGCAGGAGGTACACACAGCAGCAGCAGCAGCAGCAGCAGC 2317
 QY 2239 CTCGCGCGGCGGCGGCGGCGGCTTCCTCGTGCACCAAGCAGCAGCAGCAGCAGCAGCAGCAGC 2298
 DB 2318 ACCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2377
 QY 2299 TCCGCGAGCTCAGCAGCA 2317
 DB 2378 GAGGCTACACACAGCAGCA 2396
 RESULT 11
 MA199682
 ID AAI9682 standard; DNA: 4411529 BP.
 XX
 AC AAI9682;
 XX
 DT 15-JAN-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
 XX
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
 OS Mycobacterium tuberculosis.
 XX
 PN US6294328-B1.
 XX
 PD 25-SEP-2001.
 XX
 PF 24-JUN-1998; 98US-0103840.
 XX
 PR 24-JUN-1998; 98US-0103840.
 XX
 PA (GENO-) INST GENOMIC RES.

```

XX Fleischmann RD, White OR, Fraser CM, Venter JC;
PI WPI: 2001-647261/74.
DR
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ
XX
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (A199683) and
CC H37Rv (A199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.
XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 144602 G; 758379 T; 0 other;
XX
Query Match 2.8%; Score 69.4; DB 22; Length 4411529;
Best Local Similarity 43.2%; Pred.No. 0.00015;
Matches 503; Conservative 0; Mismatches 646; Indels 16; Gaps 3;
XX
QY 1327 GCAGAGCGCATGAGGAATTCCTCCAGAGCATGCTCCAGACCGCGGGAAGCAG 1386
DB 335004 GCGCGGAGCGCCACGATCCCGCGGAGAGCCCTGCGCACTCCGCTTCCGCGGC 335003
QY 1387 CTAATGCGCCGAGTGAGAGCAACAAGGCTTACCTTCCAGAGCAAGAACGCTCGGG 1446
DB 335064 GCGCGGAGCGCAAGACGCTCCGCGCTTCCGCGGCGCGCTTCCGCGCGCGC 335123
QY 1447 ACATGCGCGCAAGCGTCTGGGCAACGCCCATATGGCGCCATCGACGAGCGTCACTAC 1506
DB 335124 GAAGCGGAATCCGCGCGCGCGCGCGCGCGCGCGAGCGGAGAGAGCAAGCCAGCGTTGCC 335183
QY 1507 GCGGGAACCTCGGTCAAGAGCGTCTCCGCTCGGCAACCGCGCGAGCAAGAGTG 1566
DB 335184 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335243
QY 1567 GCCATCTGCTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1626
DB 335244 TCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335303
QY 1627 AAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1686
DB 335304 GTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335363
QY 1687 TCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1746
DB 335364 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335423
QY 1747 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1806
DB 335424 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335483
QY 1807 ACCCTTATCGAGCAAGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1866
DB 335484 GAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335543
QY 1867 GCAGGAGCATCCGCGGATTCGCTGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1926
DB 335544 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335603
QY 1927 ACCGACGTGCGCATATTCGCGCTCCACAGACTCGGCGCGCGCGCGCGCGCGCGCGCG 1986

```

```

DB 335604 TCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335663
QY 1987 GCCCTGACCAACACCTACAGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2046
DB 335664 GCGC---GTGAGGCGCTACTACGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335720
QY 2047 TATGCGCTTGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2106
DB 335721 CATGGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335780
QY 2107 TCCGAGCGGACATCCAGGCGCTCCAGGCGCTCCATGCGACAGCAAGGTCGCC 2166
DB 335781 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335840
QY 2167 GCGAGGCGGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2226
DB 335841 GTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335900
QY 2227 GCTCAGGAGAACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2276
DB 335901 GATTTAACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 335960
QY 2277 GCAGAGCAGCAGCAGTACTCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2336
DB 335961 GCCGATCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336020
QY 2337 ATCCACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2393
DB 336021 GCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336080
QY 2394 CACGCGGCTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2453
DB 336081 GAAGTGGGATTAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336140
QY 2454 GACGCAAGTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2478
DB 336141 GCCGAACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336165
XX
RESULT 12
AAH74540
ID AAH74540 standard; cDNA; 1107 BP.
XX
AC AAH74540;
XX
DT 15-OCT-2001 (first entry)
XX
DE Nucleotide sequence of an endochitinase polypeptide.
XX
KW Chitinobiosidase; chitinolytic enzyme; early flowering; ss.
XX
OS Streptomyces albidoflavus.
XX
PN WO200146387-A1.
XX
PD 28-JUN-2001.
XX
PF 22-DEC-2000; 2000WO-US35238.
XX
PR 23-DEC-1999; 99US-0172003.
XX
PA (CORR ) CORNELL RES FOUND INC.
XX
PI Broadway RM, Gongora CE;
XX
DR WPI: 2001-496661/54.
XX
PT Promoting early flowering, yield in plants and reducing plant size,
PT involves growing transgenic plant or plant seed transformed with DNA
PT molecule encoding chitinolytic enzyme, or applying the enzyme to the
PT plant
XX

```

PS Disclosure: Page 18-19; 87pp; English.
 XX The present sequence encodes an endochitinase polypeptide. The
 CC polypeptide is a chitinolytic enzyme. The chitinolytic polynucleotide
 CC is used to produce transgenic early flowering plants. The chitinolytic
 CC polynucleotides and polypeptides are useful for promoting early
 CC flowering in plants, promoting yield from plants and reducing plant
 CC size.
 XX
 XX Sequence 1107 BP; 202 A; 427 C; 331 G; 147 T; 0 other:
 SO
 Query Match 2.7%; Score 66.8; DB 22; Length 1107;
 Best Local Similarity 44.5%; Pred. No. 0.00013;
 Matches 405; Conservative 0; Mismatches 492; Indels 13; Gaps 3;
 1475 CCACATGGGCGACCTGACAGAGGTGACATACCGCGGAACCTGGTCAGAGCGTCTC 1534
 131 CCATCGGTGACACCTTGGCGCTTACGACAAGCGTACACCGCGCGAGTGGTGAGC 190
 1535 GCGTCGGCAACACCGCGCGACGCAACAGGTGGCATCTTCGCGAGCGCGCGACGT 1594
 191 GCGTCGGCGACACTGGAGACACCGCTGGCGGCAACTTCAACGACTCCGCAAGCTCA 250
 1595 GAGCATTCGACTAGCGGCGGACACGTCATGAAACGCGCGACGCGCTATTGCGCGC 1654
 251 AGGCAAGTACCGCGACATCAAGGTCTTGTCTTGGCGGCGGTGAGACTGGTCCGCGC 310
 1655 ACGGCGACACGATCTCTGTGACACCGCGCTGCGCGCGGTGAGCGCTGCAATTCAGG 1714
 311 GCTTCAACCGACCGCGTGAACACCGCGCGCTTGGCAAGTCTCCACGACCTGGTGC 370
 1715 GCAGCTTTGCGCTCGCTCGAGCGCTGCGCGCGCGCGCTATCGCTCGACAGAAGA 1774
 371 AGAACCGCGCGCTGGCGGACGCTTTCGACCGCATGACCTCGACTGGAGTACCGAAGC 430
 1775 CCACAGCGCTTCTTACGCGCGCTCCGAGTCACTTTTACGTCAGACAGACCGGCA 1834
 431 CCGCGCGCGCTGACGAGCTCGGCTCGCGCGCGCGTGAAGAACATGCTCAGGCGA 490
 1835 GCAGCTTCAACGCGCGCGCGCGACGCGCGCGCGCGAGATCCGCGAGATGCTGCTC 1894
 491 TCGCGCGCGCGCTGGCGACCGCGCTGCTGACCGCGCGCGCGCGCGCGCGCGCGC 550
 1895 ACCGCAACCGCGCGCGCGCGCGCTGATGTCGACGAGCGTGGCAATTCGCTCCACAG 1954
 551 GCGCAAGCTCGACGCGCGCGCGCGCTAGC--CGGCGCGCGCGCGCGCTGACTGGTACA 607
 1955 ACTGCGGCGACGACTTTGGCGCAAGTCTCCACGCGCGCTGACCAACACTTACAGATGCCC 2014
 608 ACGTGTGACGAGTCTTCTTGGCGCGCTGCGGACAAGACCGCGCGCGCGCGCGCGC 667
 2015 TGGGTGCGGCGTCAAGCTGCAAGCTGATGCTTGGCGCGCGCGCGCGCGCGCGCGC 2074
 668 GCGCGCTGAACTCTTACAGCGCGCATCCCAAGCGCGCTTCTTCCATGCGCGCGCGCGC 2134
 2075 CTGCGCTTACGCGCGTGAAGACAGCGCGCGCTCTGAGAGGACATTCAGAGCTCCAGG 2191
 728 CCAAGCTCAAGGGAAGGCGCGTCCGCGGCAAGGCTCTGCGGATGCGGCTTCTACG 787
 2135 GCTTGGGCTTCACTGACAGACGAGTCCGCGGACCGCGCGCGCGCGCGCGCGCGCGC 847
 788 GCGCGCGCGTGAAGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 847
 2192 -----TTACGTGGGCGACCAACGCGCGCGCGCTTTTACGCGTCAAGGAACGCGCGCG 2245
 848 CCGGCACTACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 907
 2246 GCGGCACTACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 907
 908 CCGGCACTACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 967
 2306 GCTGAGACACGAGCTGAGGTGCA-GCTTGTATCCACGAGCGCGCGCGCGCGCGCGCGC 2364

DB 968 CCGCGCGCACATCAAGACCAAGATGACCTGGCGCAAGGACGAGCGCTCGCGCGCGCT 1027
 QY 2365 TCGTGGAGA 2374
 DB 1028 TCTTCTGGGA 1037
 RESULT 13
 AA206825
 ID AA206825 standard; DNA; 2712 BP.
 XX
 XX AA206825;
 AC
 XX 09-NOV-1999 (first entry)
 DT
 XX
 DE Streptomyces albidoflavus endochitinase DNA.
 KW Chitin; chitinolytic enzyme; insect; fungus; resistance;
 KM transgenic plant; insecticide; fungicide; ds.
 XX
 OS Streptomyces albidoflavus.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 1535..2662
 FT /tag= a
 FT /product= "Mature endochitinase"
 FT /function= "Randomly cleaves chitin molecule internally"
 XX
 PN W09942594-A1.
 XX
 PD 26-AUG-1999.
 XX
 PF 12-FEB-1999; 99MO-US03176.
 XX
 PR 18-FEB-1998; 98US-0025691.
 XX
 PA (CORR) CORNELL RES FOUND INC.
 PI Broadway RM, Harman GE;
 PI
 DR WPI: 1999-527480/44.
 DR P-PSDB: AAT33722.
 XX
 PT New nucleic acid encoding chitinolytic enzyme active under alkaline
 PT conditions, used to generate plants resistant to insects and fungi
 PT and to produce recombinant enzyme
 XX
 PS Claim 11: Page 13-15; 90pp; English.
 XX
 CC This sequence represents Streptomyces albidoflavus endochitinase
 CC DNA. Endochitinase hydrolyses chitin, an insoluble linear
 CC beta-1,4-linked polymer of N-acetyl-beta-D-glucosamine found in all
 CC arthropods, yeast, most fungi, and some stages of nematodes. There are
 CC three types of chitinolytic enzyme activity: glucosaminidase, which
 CC cleaves monomeric units from the terminal end of chitin; chitinobiosidase,
 CC which cleaves dimeric units from chitin terminal ends; and
 CC endochitinase, which randomly cleaves the chitin molecule internally.
 CC Endochitinase and chitinobiosidase (AAV33721) from S. albidoflavus are
 CC active under alkaline conditions. This makes them particularly effective
 CC in controlling insects, as these enzymes can be ingested by insects and
 CC can then attack them by degrading their chitin-containing, alkaline
 CC digestive tracts. These chitinolytic enzymes can be applied directly to a
 CC plant to act as an insecticide or fungicide, or non-pathogenic bacteria
 CC transformed with the DNA encoding either or both enzymes can be applied
 CC to a plant. Alternatively, transgenic plants can be produced which
 CC express either or both enzymes in some or all of their tissues. A wide
 CC variety of commercially important crops could benefit from use of these
 CC enzymes e.g. sugar cane, potato, cotton and coffee.
 XX
 SO Sequence 2712 BP; 459 A; 1094 C; 830 G; 329 T; 0 other:
 Query Match 2.7%; Score 66.8; DB 20; Length 2712;
 Best Local Similarity 44.5%; Pred. No. 0.00015;

Matches 405: Conservative 0: Mismatches 492: Indels 13: Gaps 3:			
QY	1475	CCACATGGGCGCACCTCGACGAGCGTCTACGCGCGGAACCTGGTCAAGAGCGTCTCC	1534
DB	1698	CCATGGGTACAGCTTCCGCGCTTACGACAGCGCTACACCGCGCGGTGGTGACG	1757
QY	1535	GGGTGGGCAACCGCGCGCACGCAACAGGTGGCCATCTCTGTCGACGGCGCGACGT	1594
DB	1758	GGGTGGGCGACACCTGGGACAGCGCGCTCGGGGCACTTCAACAGCTCCGCAAGCTCA	1817
QY	1595	GGAGCATGCACTACCGCGCGCACAGCTCCATGACGGCGCGACGGTGGCTATTGGCGG	1654
DB	1818	AGGCCAAGTACCGGACATCAAGTCTCTGCTCTCTGCTGCGCGCTGAGCTGTGCGGCG	1877
QY	1655	ACGGGACACGATCTCTGTGAGCGCGCTGTCGCGCGCTGAGCGCTGCGATTCCAG	1714
DB	1878	GCTTACGACCGCGCTGGAAGAACCGCGCGCTTGGCCAACTCTGACAGCTGTGTG	1937
QY	1715	GGAGCTTTGCGCTCTCTGAGCGCTGCGCGCGCGCGCTCATCGCTCGGACAGAGA	1774
DB	1938	AGGACCGCGCTGGCGCGAGCTTCTGAGCGGATGACCTCGACTGAGGAGTACCGGAAG	1997
QY	1775	CCACAGCGCTCTTACGCGCGCTCGGATCGACCTTTTACGTACAGACAGACCGGCA	1834
DB	1998	CCTGGGCTTCACTGCTGCGACAGCTCGGCTCGGCGCGCTGAAGAACATGTCTCAGGCG	2057
QY	1835	GCAGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1894
DB	2058	TGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2117
QY	1895	ACCGGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1954
DB	2118	GGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2174
QY	1955	ACTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2014
DB	2175	AGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2234
QY	2015	TGGGCTGGGCTGAGGCTCGAAGCTGGAACCTGATGCTTGGCGACCGCGCGCGCG	2074
DB	2235	CGGCGGCTGAGCTCTTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2294
QY	2075	CTGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2134
DB	2295	CCAGGCTCAAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	2354
QY	2135	GCTTGGGCTTCCATCGACGACGACGACGACGACGACGACGACGACGACGACGACG	2191
DB	2355	GGCGGCGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2414
QY	2192	-----TCTAGCTGGGCGACCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2245
DB	2415	CGGGGACCTTACGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2474
QY	2246	GGGCGACGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2305
DB	2475	CGGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2534
QY	2306	GCTTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG	2364
DB	2535	CGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2594
QY	2365	TGCTGCGAGA 2374	
DB	2595	TCTTCTGGGA 2604	

RESULT 14
AAH74538
ID AAH74538 standard: DNA; 2712 BP.
XX
AC
XX AAH74538:
XX

DT	15-OCT-2001 (first entry)
XX	
DE	Nucleotide sequence of an endochitinase polypeptide.
XX	
KW	Endochitinase; chitinolytic enzyme; early flowering; ss.
XX	
OS	Streptomyces albidoflavus.
XX	
FH	Key Location/Qualifiers
FT	CDS 1535..2665
FT	FT /tag= a
FT	/product= "Endochitinase"
XX	
PN	WO200146387-A1.
XX	
PD	28-JUN-2001.
XX	
PF	22-DEC-2000; 2000WO-US35238.
XX	
PR	23-DEC-1999; 99US-0172003.
XX	
PA	(CORR) CORNELL RES FOUND INC.
XX	
PI	Broadway RM, Gongora CE;
XX	
DR	WPI: 2001-496661/54.
XX	
PT	P-PSDB: AM63553.
XX	
PT	Promoting early flowering, yield in plants and reducing plant size,
XX	
PT	involves growing transgenic plant or plant seed transformed with DNA
XX	
PT	molecule encoding chitinolytic enzyme, or applying the enzyme to the
XX	
PT	plant
XX	
PS	Claim 11: Page 15-16; 87bp: English.
XX	
CC	The present sequence encodes an endochitinase polypeptide. The
XX	
CC	polypeptide has a molecular mass of 45 kD and an isoelectric point
XX	
CC	of less than 6.5. The polypeptide is a chitinolytic enzyme. The
XX	
CC	chitinolytic polynucleotide is used to produce transgenic early
XX	
CC	flowering plants. The chitinolytic polynucleotides and polypeptides
XX	
CC	are useful for promoting early flowering in plants, promoting yield
XX	
CC	from plants and reducing plant size.
XX	
SQ	Sequence 2712 BP; 459 A; 1094 C; 830 G; 329 T; 0 other;
XX	
Query Match 2.7%; Score 66.8; DB 22: Length 2712;	
Best Local Similarity 44.5%; Pred. No. 0.00015;	
Matches 405: Conservative 0: Mismatches 492: Indels 13: Gaps 3:	
QY	1475 CCACATGGGCGCACCTCGACGAGCGTCTACGCGCGGAACCTGGTCAAGAGCGTCTCC
DB	1698 CCATGGGTACAGCTTCCGCGCTTACGACAGCGCTACACCGCGCGGTGGTGACG
QY	1535 GGTGGGCAACCGCGCGCACGCAACAGGTGGCCATCTCTGTCGACGGCGCGACGT
DB	1758 GGTGGGCGACACCTGGGACAGCGCGCTCGGGGCACTTCAACAGCTCCGCAAGCTCA
QY	1595 GGAGCATGCACTACCGCGCGCACAGCTCCATGACGGCGCGACGGTGGCTATTGGCGG
DB	1818 AGGCCAAGTACCGGACATCAAGTCTCTGCTCTCTGCTGCGCGCTGAGCTGTGCGGCG
QY	1655 ACGGGACACGATCTCTGTGAGCGCGCTGTCGCGCGCTGAGCGCTGCGATTCCAG
DB	1878 GCTTACGACCGCGCTGGAAGAACCGCGCGCTTGGCCAACTCTGACAGCTGTGTG
QY	1715 GGAGCTTTGCGCTCTCTGAGCGCTGCGCGCGCGCGCGCTCATCGCTCGGACAGAGA
DB	1938 AGGACCGCGCTGGCGCGAGCTTCTGAGCGGATGACCTCGACTGAGGAGTACCGGAAG
QY	1775 CCACAGCGCTCTTACGCGCGCTCGGATCGACCTTTTACGTACAGACAGACCGGCA
DB	1998 CCTGGGCTTCACTGCTGCGACAGCTCGGCTCGGCGCGCGCTGAAGAACATGTCTCAGGCG

Db 789 CAGCAGCACCGTCGGCTACCCGCGCAGTACCGGAGCGTCATCGCGCGCGGTGGA 848
Oy 2034 GAACGTGAACCTGTATGCGCTTGGACCGGCCGTCAGGGGCTGCGCTCTAGCCAGTGG 2093
Db 849 CAGCAGCAACCAAGCGCGGAGCTTTCAGCAGC--GTCGGCCCGAGCTGAGCTCATGGC 905
Oy 2094 AGACAGCGCGCGCTCTGAGACGACATCCAGGGCTTCCAGGGCTTCGCTCCATCGACAG 2153
Db 906 CCGGGGCTCAGCATTCAGAGCACCCTCCCGGCAACAAGTACGGCGCCCTACAGCGGCAC 965
Oy 2154 CACCAAGTCCCGGCGAGCGGAGCAGCCCGGCAAGTCTACGTGGGCACCAAC 2208
Db 966 CAGCATGGCCAGCCCGCAGCTGCGCGCGCTGCTCATCTCATGCAAGCAC 1020

Search completed: May 4, 2003, 15:04:23
Job time : 15130.7 secs

THIS PAGE BLANK (USPTO)

C	1	88	3.5	4403765	4	US-09-103-840A-2	Sequence 2, Appl
C	2	76.8	3.1	1833	2	US-08-403-852D-6	Sequence 6, Appl
C	3	76.8	3.1	1833	3	US-08-510-646B-6	Sequence 6, Appl
C	4	76.8	3.1	1833	4	US-09-231-818-6	Sequence 6, Appl
C	5	71.4	2.8	4403765	4	US-09-103-840A-2	Sequence 2, Appl
C	6	70	2.8	2793	1	US-08-209-747-1	Sequence 1, Appl
C	7	70	2.8	2793	1	US-08-458-298-1	Sequence 1, Appl
C	8	69.4	2.8	441529	4	US-09-103-840A-1	Sequence 1, Appl
C	9	66.8	2.7	2712	3	US-09-023-691-4	Sequence 4, Appl
C	10	65.4	2.6	1140	3	US-09-023-173-4	Sequence 4, Appl
C	11	60	2.4	2810	4	US-09-408-647A-1	Sequence 1, Appl
C	12	60	2.4	441529	4	US-09-103-840A-1	Sequence 1, Appl
C	13	59.6	2.4	2367	4	US-09-056-556-201	Sequence 201, App
C	14	59.6	2.4	2367	4	US-09-072-596-196	Sequence 196, App
C	15	58.4	2.3	2249	3	US-08-814-052-19	Sequence 19, Appl
C	16	58.2	2.3	1656	4	US-09-385-028-14	Sequence 14, Appl
C	17	58.2	2.3	15079	4	US-09-385-028-1	Sequence 1, Appl
C	18	57.4	2.3	2064	1	US-08-343-428-1	Sequence 1, Appl
C	19	56	2.2	1248	4	US-09-105-537-7	Sequence 7, Appl
C	20	56	2.2	5970	3	US-09-320-878-21	Sequence 21, Appl
C	21	56	2.2	13613	3	US-09-105-537-3	Sequence 3, Appl
C	22	55.8	2.2	3332	2	US-08-403-852D-1	Sequence 1, Appl
C	23	55.8	2.2	5392	3	US-08-510-646B-1	Sequence 1, Appl
C	24	55.8	2.2	5392	4	US-09-231-818-1	Sequence 1, Appl
C	25	55.8	2.2	68750	3	US-09-335-409-1	Sequence 1, Appl
C	26	55.8	2.2	68750	4	US-09-568-102-1	Sequence 1, Appl
C	27	55.8	2.2	68750	4	US-09-567-965-1	Sequence 1, Appl

C 28	55.8	2.2	687.50	4	US-09-568-480-1	Sequence 1, Appl. 1
C 29	55.8	2.2	687.50	4	US-09-568-480-1	Sequence 1, Appl. 1
C 30	55.8	2.2	687.50	4	US-09-568-480-1	Sequence 1, Appl. 1
C 31	55.8	2.2	687.50	4	US-09-567-899-1	Sequence 1, Appl. 1
C 32	55.8	2.2	71989	4	US-09-443-501A-2	Sequence 2, Appl. 1
C 33	55.4	2.2	2214	3	US-08-864-038A-1	Sequence 1, Appl. 1
C 34	55.4	2.2	3311	3	US-08-864-038A-2	Sequence 2, Appl. 1
C 35	55.4	2.2	3311	3	US-08-864-038A-4	Sequence 4, Appl. 1
C 36	55	2.2	1896	3	US-09-311-626B-15	Sequence 15, Appl. 1
C 37	55	2.2	3300	4	US-08-194-290-6	Sequence 6, Appl. 1
C 38	55	2.2	3300	2	US-08-614-377A-6	Sequence 6, Appl. 1
C 39	55	2.2	3300	4	US-09-142-648B-6	Sequence 6, Appl. 1
C 40	54.8	2.2	28958	1	US-08-258-261B-6	Sequence 6, Appl. 1
C 41	54.8	2.2	28958	1	US-08-456-837-6	Sequence 6, Appl. 1
C 42	54.8	2.2	28958	1	US-08-457-342-6	Sequence 6, Appl. 1
C 43	54.8	2.2	28958	1	US-08-457-646A-6	Sequence 6, Appl. 1
C 44	54.8	2.2	28958	1	US-08-458-076A-6	Sequence 6, Appl. 1
C 45	54.8	2.2	28958	1	US-08-764-233A-6	Sequence 4, Appl. 1

ALIGNMENTS

```

RESULT 1
US-09-103-840A-2/c
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: TITLE OF INVENTION: TUBERCULOSIS
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequences
: OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

Query Match:	3.5%;	Score 88;	DB 4;	Length 4403765;
Best Local Similarity	43.0%;	Pred. No. 4.9e-10;		
Matches 433;	Conservative	0;	Mismatches 575;	Indels 0;

OY 1391 TGCGAGGTGCGAACAACAAAGGGCTTCACCTTTGCCAGCAAGAAAACGCATCTGGGAACAT 1450
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3937116 TCGCCGCCGTCGCCGCCGCTCCGCCGCTCACCCTCCCGGCCGCCCAACACACCG 3937051

OY 1451 CGCGCGAGAAGGTCTTGSGAACGCCCCACATGGGGCACCTTGCAGAGAGCTTCAGACTPACGCC 1510
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3937056 CCACCGTTTACCGCCTTGTGCCACCGAGCGCGGGGAGCTGTGCCACCGCCACAGCGGTGCG 3936997

OY 1511 GGAACTCGGTTCAAAGAGCGTCGTCCCCTGTGCGGAACACGCCCGGCAACGCAAMCAGGTGGCCA 1570
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3936996 TCGCCCCCGGTGCGCGCGCGGACACGGCGCTGTGCGCGCGGCCAACCTTAGACCCCCGGTCCCG 3936933

OY 1571 TCTCGTCCGACGGCGGGCGAGCTGAGACATGACTACCGGGCCGACACGTCATTGAAACG 1630
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3936936 CCAACGGGTGCGCTCTTTGGCCGCCGCCGCAACCTTTCGGCGGGCGCCCGCTTAACCGCGCG 3936877

OY 1631 GCGGACGGTGGCATTTATTCGGCGSACGGGAGACACGATCTCTGTGTGCAGCGCTGTGTCGG 1690
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3936876 GCCCGCGCGGTGCGCGCGGTCGCGCGCGGTATATTGCGCGCGCTTGCGCGCGCGGTCCCGCCG 3936811

QY 1691 GCGTCAAGCGCTTCGAGATTCAGAGGCGAGCTTTCCTCCGCTTCAGAGCTGCCCCGGGGCG 1750
DB 3936816 GCACCGCGCGGACACACCTTGACACACCGGACCGCGCGCGCGCGCTTGCGACACTTGT 3936757
QY 1751 CGGTATCGCTTCGAGCAAGAGACCAAGAGCTTCTTACGCCGCTTCGGATGACCT 1810
DB 3936756 CCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCGCGTCCGCGCGCGCGCGCG 3936697
QY 1811 TTACCTAGCAAGAGACACCGGACAGCTTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1870
DB 3936696 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936637
QY 1871 GGACGATCCGGGATATCGCTGCTACCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1930
DB 3936636 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936577
QY 1931 AGCTGCGCATATTCGCGCTTCACAGACTTCGGGACACGCTTTGGCCAAAGTCTCACCGCGCG 1990
DB 3936576 TTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936517
QY 1991 TGACCAACACCTACAGATCGCCCTGGGTGTGGGCTCAGGCTCAGCACTGGAACCTGTATG 2050
DB 3936516 TGACCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936457
QY 2051 CCTTGGGACCGCGCGCGCTCAGGGGCTCGCTCTACGCGCAGTGAGACAGCGCGCGCTCT 2110
DB 3936456 CCGCGACCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936397
QY 2111 GGACGACATTCAGGCGCTCCACAGGCGCTTCGCGCTTCAGACAGACCAAGAGCGCGCGCA 2170
DB 3936396 ATACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936337
QY 2171 GCGGACAGCACCGCGCGCGCAAGTCTACGTGGGACCAACGCGCGCGCGCGCTTTTACGCTC 2230
DB 3936336 GCGGACGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936277
QY 2231 AAGGAACCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2290
DB 3936276 TTGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936217
QY 2291 GTACCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2350
DB 3936216 CCACCGGACCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936157
QY 2351 CTTGCAAGGTGATCTGTCGAGAGACAGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2398
DB 3936156 CTACACCGCGGTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3936109

RESULT 2
US-08-403-852D-6
Sequence 6, Application US/08403852D
Patent No. 5891695
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESS: Finegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA

ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403, 852D
FILING DATE: 10-MAY-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806, 0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 103..1689
US-08-403-852D-6

Query Match 3.1%; Score 76.8; DB 2; Length 1833;
Best Local Similarity 45.9%; Pred. No. 1.8e-08;
Matches 338; Conservative 0; Mismatches 392; Indels 6; Gaps 2;

QY 1483 GCCACCTGCACGAGCTGCACATACGCGGGAACCTGGTAAAGAGCTGCTCCGCGTGGC 1542
DB 820 GCCATCATCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 879
QY 1543 AACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1602
DB 880 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 939
QY 1603 GACTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1662
DB 940 CTCGACATGAGGCTTCTTACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 996
QY 1663 AGCATCTCTGTCGACCGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1722
DB 997 TCGTCGTTGGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1056
QY 1723 GCGTCGCTCGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1782
DB 1057 GCGTACGAGCGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1116
QY 1783 GTCTTCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1842
DB 1117 CTCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1176
QY 1843 ACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1902
DB 1177 GCGTGGGCGATACACTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1236
QY 1903 ACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1962
DB 1237 CCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1296

QY	1963	ACGACCTTTGGCCAAAGTCATACCCGGCTCGACCAACACATACACATATGCCCTG---GGT	2019
Db	1297	CTGGCCATGCCCCCGCATGGCCACCGCCCTGATATGTCTCTCATTCGCCCGCCCAAGGCCGGG	1356
QY	2020	GTCGGCTCAGAGCTCGAATCGGAACCTGTATGCCCTTCGGACACGGCCCTCAGAGGGCTGC	2079
Db	1357	GCGCGAAGCGGGCTGCGAGGGACACCTCGAGACCTGAGTTCGGCGGGGAGCTGGAGTGGATC	1416
QY	2080	CTCTACGCCCATGTGGACACAGCGGCCCTCTCTGAGACGGAATCAGAGGCTCCAGAGGCTTC	2139
Db	1417	CTGCGGGCGCTCTCCGCTCCCGCTTCGCTTCCCAACTGCCCCGCCCATATCCGGCACCC	1476
QY	2140	GGCCTCATCTGACAGCACCAAGGTGCGCCGACAGCGGACAGCACCGCCGGGCAATGTACATGTG	2199
Db	1477	GGCTCCCTCGACAGGACATCGCCCGACGCGCACACCCAAACAGCGCGGCGAGTCCACAGAC	1536
QY	2200	GGACACCAAGGCCGGG	2215
Db	1537	GCGTTGCGCGACGCGG	1552

RESULT 3

US-08-510-646B-6

```

; Sequence 6, Application US/08510646B
; Patent No. 6077699

```

; GENERAL INFORMATION:

1 APPLICANT: Blanc, Veronique
 2 APPLICANT: Blanche, Francine
 3 APPLICANT: Crouzet, Joel
 4 APPLICANT: Jacques, Nathalie
 5 APPLICANT: Lacroix, Patricia
 6 APPLICANT: Thibaut, Denis
 7 APPLICANT: Zagorec, Monique
 8 APPLICANT: Debussche, Laurent
 9 APPLICANT: De Creely-lagard, Valerie
 10 TITLE OF INVENTION: Polypeptides Involved In The
 11 Biosynthesis Of Streptogramins, Nucleoside Sequences
 12 TITLE OF INVENTION: Coding For These Polypeptides And Their Use
 13 NUMBER OF SEQUENCES: 45
 14 CORRESPONDENCE ADDRESS:

COMPUTER READABLE FORM;

```

: MEDIM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:

```

; PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993

; PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 92/11441

ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.

REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 02

REFERENCE/DOCKET NUMBER: 03806.0054-010000
TELECOMMUNICATION INFORMATION:
; ;
; ;

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

; INFORMATION FOR SEQ ID NO: 6:

```

?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 1833 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: double
?      TOPOLOGY: linear
?      MOLECULE TYPE: cDNA
?      HYPOTHETICAL: NO
?      ANTI-SENSE: NO
?      ORIGINAL SOURCE:
?      ORGANISM: S.pristinaespiralis
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 103..1689
?      US-08-510-6468-6

```

Query Match	3.1%	Score 76.8	DB 3	length 1833	.
Best Local Similarity	45.9%	Pred. No. 1.8e-08			
Matches 338	Conservative 0	Mismatches 392	Indels 6	Gaps 2	

QY	1483	GCACCTCGACGACGACGCTGCATCGCCGGGAACTCGGTTCAGAAAGCGTCGTCGGCTGGC	1542
Db	820	GCCTATCTCTCCGGCCCCGAAACAGCGCTGAGCGGCCCCGACAGTCTCTGCGGGCCCC	879
QY	1543	AACACCCCGCGGACGACAAAGATGGCCATCTGCTCGACGCGGGCGAGCTGAGCATC	16020
Db	880	GTCGCGGCGCGCGCGCTCACCGCGTTGTCGCTGGGAATCGACACCCCAACCCCATG	939
QY	1603	GACTACGCGGCGGACACGTCATGAAACGGCGGACCGTGGCTATTGCGCGGACGGCAC	16620
Db	940	CTCCACATGGGCTTCTTTCACCGACCGGCGCTTCAAGGGGCGCTTCCCGGCGAGTGC	996
QY	1663	ACGATCTCTGTGTGACACCGGCTGTGCGGGCGTGCAGCGCTGCAATCCAGGCGAGCTT	17220
Db	997	TCTGTCGTTCGGCATGGCCCGGCTGCTTTCCTGCTGCTACCCAGCACTCCAACTGCTTC	10560
QY	1723	GCCCTCCCTCGAGCGCTGCCCCGGGGCGGCCGTCATCGCTTCGAGCAAGAACACACAGC	17820
Db	1057	GGCTACAGACGCCCTGCAAGCGCGGCGCTGTGGACCGGCGCATGCGTTTGACGATGCTGCC	11160
QY	1783	GTCCTTACGCGCGGCTCCGAGTGCACCTTTTACGTGACGAAAGACCGGCGACGCTTC	18420
Db	1117	CTCAACCTGGCGCGGCTCGGCGGAAACTCTGCGCGCTCGGACACGCGCGCGAGCATC	11760
QY	1843	ACGGGGGGGGCCAAAGGTGGGCGACGGAGGAGAGATCCGGGATATGCTGCTACCCGAC	19020
Db	1177	GCCTTGGGCGATGACACTGCTGGCGCGCGGCTCTAGCGCGGTGGCTCGCGGATGGGC	12360
QY	1903	ACCGCGGGACAGTGTGATGCTGTCGACCGAGCGTGGCATTTTCCGCTCACAGACTCGGGC	19620
Db	1237	CCCGACGCGGGCTACGGCGGCAAGTCTCGCGCGGCTGCTCTCATATGGGCGGCGCATGCCA	12660
QY	1963	ACGACCTTTGGCCAAAGTCTCCACCGCGCTGACCAACACCTACAGATGCGCCTG--GGT	20190
Db	1297	CTGGCGCATGGCCGCCATGGCGACCGCGCGTGAATGTCCTTCATCCCGCGCAAGCGCGGG	13560
QY	2020	GTTGGGCTCAGGCTGCAACTGGAACCTGTATGCTTGTGGACGCGCGCGCTCAGGGGCTGCG	20790
Db	1357	GCGCGAAGCGGGCGTGCAGGGCACCTGACCGAGTGTGGGGGGGAGCTGGGAATGGCGATC	14160
QY	2080	CTTACGCGCAGTGGAGACAGCGCGGCTCTCTGAGACGGATCCAGAGGCTCCGAGGGCTTC	21390
Db	1417	CTGGGCGCGCTCTCGGCTTCCGCTTTCGCTTCCCAACTCCCGCGCATCACCGGCAAC	14760
QY	2140	GCGTCATGACAGACCAAGGTGCGCCGCGAGCGGACGACCGCGCGCAAGTTCATGATG	21990
Db	1477	GCGTCCCTTCACAGAGGACATGCGCGACGCCACACCCACAGAGCGCGGAGGTCCACGAC	15360
QY	2200	GGCACAACGGCGCGG 2215	
Db	1537	GCCTTCCGCGAGCGG 1552	

RESULT 4

[illegible]

Db 880 GTCCGGGCGCCGGCGCTACCGCGCTTGTCGCGCTGGGAATGCAACACCCCCACCCCAATG 939

QY 1603 GACTACGCGGCCGACACAGCTCCATGACACGGCGGCACAGTGGCCCTATTTCGGCGCACGGCGAC 1662

Db 940 CTTCACATATGGGCTTCTTCCACCGACCGGCGCTTCAACAGGGCGCTGTGGCGGGAGTG--C 996

QY 1663 ACGATTCCTGTGTGACACCGCGCTGTCGGGGGTGACGAGGTCTGCGAGATTTCAGGGGACGCTTT 1722

Db 997 TCGTCTGTTCCGACATGACGCGCGGTGCGCTTCTCTGTCTACCCAGACCTCCAACTCTGTCTTC 1056

QY 1723 GCGTCCGTTCTGACGCTGCCCCGCGGGCGCGCTCATCGCTCTGGACAAGAACCAACAGC 1782

Db 1057 GGCTATACGACGCCCTCGACAGCGCGGCGCTCGACACGCGCGCACCTGGCTTTGAGATCTGTGCGC 1116

QY 1783 GTCCTTACCGCGGCTCCGGATATGACCTTTTACGTACGAAAGACACCGGACACGCTTC 1842

Db 1117 CTCAACCTGCGCGGCGCTCGGCGGAAACTCTTCGCGCGGCTCTCGGACACGCGCGCACATTC 1176

QY 1843 ACGCGCGGCGCCAGCTGTGGCGAGCGACGAGGACGATCCGGGATATTCGCTCACCGCGACC 1902

Db 1177 GCGCTGGGACATGACATCTGTGGCGCGCGCGGCGCTCAGCGCGGTGGCGCGGCGGATGGGCG 1236

QY 1903 ACCGCGGGCGACGTTGTATGTCTGACCGACGATGTGGCATATTTCGCTCCACAGACTCGGCG 1962

Db 1237 CCGGACGCGGCGCTACGCGCGGCGATCTCGCGGCGCTGCTCTCATGGGCGGGGCGATATGCA 1296

QY 1963 ACGACCTTTGGCCCAAGTCTCCACCGCGCTTACCAACACCTACCAAGATCGCGCT---GGT 2019

Db 1297 CTGGGCAATGCCCGCATGGCGCACCGCGCGGTATGTCTTCATCCCGCGCGCAAGGCGCGG 1356

QY 2020 GTGGGCTACAGGCTCTGAACCTGAACTTATAGCTTTCGGCACCGCGCGCTGACGGGCGCTCG 2079

Db 1357 GCGCGAGCGGGCGCTGACAGGGCAACCTTACCGAGATTTCGGCGCGGACTGGGAGTGGGATC 1416

QY 2080 CTCTACGCGCACTGGAGACAGCGGCGCTCTCTGACGGACATTCAGAGGCTCCACGGGCTTC 2139

Db 1417 CTGCGGCGCGCTCTCTGCGCTCCGCTTCGCTCCCACTGCGCGCGCATCACCGGACAC 1476

QY 2140 GCGTCATCTGACACACCAAGATGTCGGCGGAGCGGACAGCACCGCGGGGCAAGTCACTGTC 2199

Db 1477 GCGTCCCTGCGAGGACATGCGGAGCGCACACCCCAACAGAGCGCGGCGAGTCCACGAC 1536

QY 2200 GGCACCAACGCGCGCGG 2215

Db 1537 GCGTTCGCGACGGCG 1552

RESULT 5

US-09-103-840A-2

: Sequence 2, Application US/09103840A

: Patent No. 6294328

: GENERAL INFORMATION:

: APPLICANT: FLEISCHMAN, Robert D.

: APPLICANT: WHITE, Owen R.

: APPLICANT: FRASER, Claire M.

: APPLICANT: VENTER, John C.

: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

: TITLE OF INVENTION: TUBERCULOSIS

: FILE REFERENCE: 24366-20007.00

: CURRENT APPLICATION NUMBER: US/09/103,840A

: NUMBER OF SEQ ID NOS: 2

: SOFTWARE: PatentIn Ver. 2.1

: SEQ ID NO 2

: LENGTH: 4403765

: TYPE: DNA

: ORGANISM: Mycobacterium tuberculosis

: FEATURE:

: OTHER INFORMATION: CDC 1551

: OTHER INFORMATION: "n" bases at various positions throughout the sequence

: OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2


```

Db 1102 GCACGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1043
Oy 1664 CGATCTCTGTGTGAGCAGCCTCTGTCGGGGGTGAGCGCTGCAGTTTCCAGGAGCTTGG 1723
Db 1042 GCACGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1111
Oy 1724 CTTCTCTGTGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1783
Db 982 GCACGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 923
Oy 1784 TCTTCTACGCGGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1843
Db 922 GCACGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 863
Oy 1844 CGCGGCGGCGGCGGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1903
Db 862 GATCTCTGTGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 803
Oy 1904 CGCGGCGGCGGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1963
Db 802 TGACGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 743
Oy 1964 CGACCTTTGGGCGGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 2023
Db 742 CTACGCTTAACCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 683
Oy 2024 GCTCAGCTTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 2083
Db 682 GCTCAGCTTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 623
Oy 2084 ACAGGAGTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 2143
Db 622 CCAGGAGTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 563
Oy 2144 CCATGAGTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 2203
Db 562 CCATGAGTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 503
Oy 2204 CCAAGGAGTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 2263
Db 502 CCGTACGCTTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 443
Oy 2264 CCGTACGCTTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 2275
Db 442 GCTCAGCTTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 431

```

```

RESULT 7
US-08-458-298-1/c
; Sequence 1, Application US/08458298
; Patent No. 575677
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; TITLE OF INVENTION: CDNs Encoding Minor Ampullate Spider
; TITLE OF INVENTION: Silk Proteins
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,298
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 530

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/209,747
; FILING DATE: 14-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2793 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Nephila clavipes
; TISSUE TYPE: minor ampullate gland
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 183..2675
; OTHER INFORMATION: /product="N. clavipes minor
; OTHER INFORMATION: ampullate silk protein"
; US-08-458-298-1

```

Query Match 2.8%; Score 70; DB 1; Length 2793;

Best Local Similarity 41.0%; Pred. No. 7e-07; Mismatches 467; Indels 0; Gaps 0;

```

Oy 1484 CCACCTTGAGAGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1543
Db 1222 CCACCTTGAGAGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1163
Oy 1544 ACACGCGGCGGCGGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1603
Db 1162 GCACGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1103
Oy 1604 ACTAGCGGCGGCGGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1663
Db 1102 GCACGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1043
Oy 1664 CGATCTCTGTGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1723
Db 1042 GCACGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 983
Oy 1724 CTTCTCTGTGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1783
Db 982 GCACGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 923
Oy 1784 TCTTCTACGCGGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1843
Db 922 GCACGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 863
Oy 1844 CGCGGCGGCGGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1903
Db 862 GATCTCTGTGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 803
Oy 1904 CGCGGCGGCGGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 1963
Db 802 TGACGAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 743
Oy 1964 CGACCTTTGGGCGGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 2023
Db 742 CTACGCTTAACCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 683
Oy 2024 GCTCAGCTTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 2083
Db 682 GCTCAGCTTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 623
Oy 2084 ACAGGAGTGAAGCTCTGCTCCGCGACCATATCCGCTTGTCCACCATATCCAGCACTGCACCT 2143

```

Db 622 CCACAGCGCGGACGAGCTCCAGACACACCTTACCTAACCCAGCCCTCCGGCT 563
OY 2144 CCATGACAGCAGCAGAGGTGCGCGGACGAGACCCCGGGAATCTACGTGGGCA 2203
Db 562 CCGCTCCAGAGCTGAGCAGCAGACACCTCCGGCTCCGGACCATATCTCTTGTGCA 503
OY 2204 CCACGCGCGGCGCTTTTACGCTCAGGAACTCCGCGCGGACGCGGAGCTT 2263
Db 502 CCGTAGCTCCAGCAGCTCTTCCGTACCTCCAGCGCGGACGCTCTGTGAGCAACA 443
OY 2264 CCGTGTGACCA 2275
Db 442 GCTCCAGCTCCA 431

RESULT 8
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529

ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query March 2.8% Score 69.4; DB 4: Length 4411529;
Best Local Similarity 43.2%; Pred. No. 7.6e-06;
Matches 503; Conservative 0; Mismatches 646; Indels 16; Gaps 3;

OY 1327 GCAGACGCGATCAGAGAAATTCCTCCGACAGACCTGCGCTGACCGCGAAGCGAG 1386
Db 335004 GCGCGACACCCACACGATCCGCGCGGACGAGCCCTGCGCACTCCGCCCTGCCCGCC 335063
OY 1387 CTATTTGCGCGAGTCGAGACGACAAAGGCTTACCTTTGCCAGAGAAAGCACTCGGG 1446
Db 335064 GCGCGCGGAGCGGAAAGACGCTGCGCGCTGCCCGCGGCGCGCTTCCGCGCGG 335123
OY 1447 ACATGCGCGAGAGGCTTGGGCAACGCCCACTGAGGCGCTGACAGAGCTGACTAC 1506
Db 335124 GAGCGCGAATCCGCGCGCGCGGAGCGCGGAGCGCGGAGGAGACACCCAGCGCTTCC 335183
OY 1507 GCGCGGAACTCGCTCAGAGAGCGCTCCGCTCGGCAACACCGCGGAGCAAGAGTGTG 1566
Db 335184 GCGCGCGCGCGCGCGCGCGCTATGCGCGCGCGCTGAGAGTACCGCGCTGCCACCGCAT 335243
OY 1567 GCCATCTCTCGAGGCGCGCGCGAGCATGACTAGACGCGCGGAGCAACGATGTCATG 1626
Db 335244 TCCCGCGCGCGCGCGCGCGCGGAGGCGAGCATGCGCGCATGTCGCGCGCGCGCC 335303
OY 1627 AAGCGCGAGAGGCTTATTTGCGCGAGCGCGAGCATGCTTGTGTCAGCCCGCTCG 1686
Db 335304 GTCCCGCGCGCGAGCAGAGCTGTGTCCCGCTGCGCTTCCCGCTGCGCGCAACAG 335363
OY 1687 TCCGCGGTGACAGCGCTCCAGTTCAGGAGCACTTGTCTCCGCTCCAGCTGCGCGCG 1746
Db 335364 CCGCGCGCGCGCGCGCGCGCGCGAGCTCCGCGGAGCTGTGCGGAGAACCCCGCATG 335423
OY 1747 GCGCGCTCATGCTCTCGAGACAGAGCAACAGCGCTTCTACGCGCGGCTCCGAGTCG 1806

Db 335424 CCGCGCGTGCCTCCGCGCGGCGGAGACGCCCGCAAGCCGACGCCAGCGGCCGCCCGCC 335483
OY 1807 ACCTTTACGTACGACAGAGACACCGGAGAGCTTACGCGGGGCGGCAAGCTGGGAGC 1866
Db 335484 GAGCTGCGCGCGCGCGCGGATCCCGGAGACCCCGCGGTGGGAGAACCGCGCGCGCC 335543
OY 1867 GCAGGAGCATCCGAGATATGCTGCTACCCGAGACACCGCGGCGAGCTTGTATGTCTG 1926
Db 335544 GCGCGCGCGCGCGCGCGCGCGGAGTACGACTGCGTGTGATGATCGCGCGCGCGCG 335603
OY 1927 ACCGACGTGCGCATATTCGCTCCAGACAGCTGCGGACAGCACTTTGGCAAGTCTCAC 1986
Db 335604 TCCGCGCGCGCGGAGAGACAGCGCGCTTGTGCGCGCACCGCGCGCGCGCGCGCC 335663
OY 1987 GCGCGGACCAACCTTACAGATGCGCGTGGGTGTGGGTCTGAGCTCAAGCTGAGACCTG 2046
Db 335664 GCGC---GTGAGGGCTACTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGAG 335720
OY 2047 TATGCTTTCGCGACCGCGCGCGCTCAGAGGCTGCGCTTACGCGAGTGAAGACAGCGCG 2106
Db 335721 CATGCGCTTGGCGCGCGCGCTCCGCGGAGACCGCGCATCCGACTGTGGGAGACCGCGCAG 335780
OY 2107 TCTTGAGAGACATCCAGAGGCTCCAGGCTTGGCTCCGCTCAGACAGCAAGAGTCCG 2166
Db 335781 GCGCGCGCGCGCGCGCTTCCGATGAGCGCGCGCGCGCGCTTGCAGCGCGCGCG 335840
OY 2167 GCGAGCGGACAGCGCGCGCGGAGTACGTTGAGGCGCACCAAGCGCGCGCGCTTTAC 2226
Db 335841 GTTGGCGCGCGCGCGCGCTTGTGAGCGCGCGCGCGCGCGCTTCCGCGCGCGCTTCC 335900
OY 2227 GCTCAGGAGACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2276
Db 335901 GATTAAACAGCGCGCGCTCCCGCGCATTTGGCCCGCGGTGCGGGGCGCGTGTGCGCG 335960
OY 2277 GCAGAGCAGACAGTACCTTCTCCGCGAGCTGAGACACAGCTGTAGAGTGCAGCTTGT 2336
Db 335961 GCGGATCAACGCGCGCGCGGATTCGCGAGAGAACTGTGTGATCGGATCCAGCAGCG 336020
OY 2337 ATCCAGACCGCGCGCTTGGAGCGTACCTTTCGCGAGAGCAACGCTGCGCGCGCG 2393
Db 336021 GCACACCGCGCGCGCTGCGCGCGCGCGCATAGCGCGCGCACCGGAGTCAATGCTTGCAC 336080
OY 2394 CACGCGGTACAGGCGCTGCGCGCGCTATATGCTACAGCGGAGGATTTGGGTGAGCGCG 2453
Db 336081 GAAGTGGGATTAAGCGCTGCGCTTGGCGCTGAGCGCGCTGATAGCGCTGCGCGCG 336140
OY 2454 GACGCGTGTGTGCGCGCTATGTC 2478
Db 336141 GCGGAAACAGCGCGCGGATGCGTGC 336165

RESULT 9
US-09-025-691-4
Sequence 4, Application US/09025691
Patent No. 6069299
GENERAL INFORMATION:
APPLICANT: Broadway, Roxanne M.
APPLICANT: Harman, Gary E.
TITLE OF INVENTION: FUNGUS AND INSECT CONTROL WITH
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/025.691
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/20120
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2712 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-025-691-4

```

Query Match Best Local Similarity 2.7%; Score 66.8; DB 3; Length 2712;

Matches 405; Conservative 0; Mismatches 492; Indels 13; Gaps 3;

```

Oy 1475 CCACATGGGCCACTCGACGAGCGTCTGACTACGCGGGAACCTCGGTCAAGAGCGTCTCC 1534
Db 1698 CCATCGGTACAGCTTCGCGCGCTACGACAGCGTACACGCGCGGAGTGCCTGACG 1757
Oy 1535 GCGTGGGCAACACCGCGGACGCAACAGTGGCCATCTCTGCTCCGAGCGGCGGAGCT 1594
Db 1758 GCGTGGGCAACACCGCGGACGCAACAGTGGCCATCTCTGCTCCGAGCGGCGGAGCT 1817
Oy 1595 GGAGCATGACTACCGCGGCGGACGCTCCATGAAGCGCGGCGGCGCTATTCGCGCG 1654
Db 1818 AGGCCAAGTACCGCGGACATCAAGTCTCTGTCTCGCGGCGGCGGAGTCTGCTCGCGG 1877
Oy 1655 AGCGGACACAGTCTCTGTGTGACGCGGCTGTCTCGCGGCGGCGGAGTCTGCAAG 1714
Db 1878 GCTTACCGAGCGCTGTGAAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1937
Oy 1715 GCAGCTTTCCTCGCTCTGACGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1774
Db 1938 AGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1997
Oy 1775 CCAACAGCGTCTTCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1834
Db 1998 CCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2057
Oy 1835 GCAGCTTTCCTCGCTCTGACGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1894
Db 2058 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2117
Oy 1895 ACCGACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1954
Db 2118 GCGGCAAGCTCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2174
Oy 1955 ACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2014
Db 2175 AGGTGATGACGACGACCTTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2234
Oy 2015 TGGGTGTGGGCTCAGGCTCGAAGCTGTAATGCTTGGGCGGCGGCGGCGGCGGCGGCGG 2074
Db 2235 CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2294
Oy 2075 CTGCGCTTACGCGGAGTGAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2134
Db 2255 CCAACTCAAGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2354
Oy 2135 GCTTGGCTCCTCAAGACAGCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2191
Db 2355 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2414
Oy 2192 -----TCTAGTGGGCAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2245

```

```

Db 2415 CCGGCACTACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2474
Oy 2246 GCGGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2305
Db 2475 CCGGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2534
Oy 2306 GCTTGGCTCCTCAAGACAGCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2364
Db 2535 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2594
Oy 2365 TCGTGGAGGA 2374
Db 2595 TCTCTGGGA 2604

```

RESULT 10

US-09-023-173-4

; Sequence 4, Application US/09023173

; Patent No. 6066781

; GENERAL INFORMATION:

; APPLICANT: Rodiguez, Thomas D.

; TITLE OF INVENTION: Production of Mature Proteins

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dehlinger & Associates

; STREET: 350 Cambridge Ave., Suite 250

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94306

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/023.173

; FILING DATE: 13-FEB-1998

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 60/038,168

; FILING DATE: 13-FEB-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Pelichory, Joanne R

; REGISTRATION NUMBER: P42995

; REFERENCE/DOCKET NUMBER: 0665-0007.30

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-324-0980

; TELEFAX: 650-324-0960

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1140 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: codon-optimized Ram3D signal-prosubtilisin Bpw'

; US-09-023-173-4

Query Match Best Local Similarity 2.6%; Score 65.4; DB 3; Length 1140;

Matches 323; Conservative 0; Mismatches 386; Indels 6; Gaps 2;

```

Oy 1494 GAGCGTCACTACGCGGGAATCGGTCAAGAGCGTCTCGCGGCGGCGGCGGCGGCGGCGG 1553
Db 312 GAGCGTCCGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 371
Oy 1554 CAGCAAGAGTGGGCAATCTCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1613
Db 372 CCGCAAGAGTGGGCAATCTCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 431

```



```
QY 1614 CGACAGTTCATGAACGGCGGACGAGTGGCTTATTCGGCCGACGGGACATCCTCTG 1673
    || || || || || || || || || || || || || || || || || || || || ||
DB 432 CAAGGTCGGCGGGGAGTGAAGCATGTTCCGAGCGACCAACCGTTCCAGAGCACAA 491
QY 1674 GTTCGACCCCTGCTCCGGCTGTCAGAGCTCCGATGTTCCAGGGCAGCTTTGCTCGTCTC 1733
    || || || || || || || || || || || || || || || || || || || || ||
DB 492 CAGCCATGACCCACAGTCCGGGACCCCTCCGCCCTCACACAGCATGAGTGGCGCTCT 551
QY 1734 GAGCTGCGCGGGGGGGCGCTGATCGCTCGGACCAAGAACCAACAGGCTTCTTGAGCC 1793
    || || || || || || || || || || || || || || || || || || || || ||
DB 552 CGGCGTCCGCGGAGCGAGCTCTAGCGCTCAAGTACTGCGGCGGACGAGCGCGG 611
QY 1794 CGGCTCGGATGACCTTTTACGTAGCAAGACACCGGACAGGTTCCACCGCGGGCGC 1853
    || || || || || || || || || || || || || || || || || || || || ||
DB 612 CCAGTACACTGATCATCAAGGCAATGATGGGCAATCGGCAACACATGAGAGTCAAT 671
QY 1854 CAACTGGGACGACGAGGACATCCGGATATCGCTCTACCCGACACCGCGGGGAC 1913
    || || || || || || || || || || || || || || || || || || || || ||
DB 672 CACCATGAGCTCGGCGGGCCGA--GCGGCAGCGCGGCTCAAGGGCGCGCTGACAA 728
QY 1914 GTTATGTCGACGACGAGTGGGATATTCGCTCCACAGACTCGGAGCAGACCTTTGG 1973
    || || || || || || || || || || || || || || || || || || || || ||
DB 729 GCGCGTCCGACGCGGCTGTGCTGTCGCGCGCGCGGCAAGAGGACGACGCGGACG 788
QY 1974 CCAAGTCTCCACCGCTTGACCAACACATCCAGATCGCCTGGGCTGAGGCTCAAGCTC 2033
    || || || || || || || || || || || || || || || || || || || || ||
DB 789 CAGCAGACACCTGCTGCTACCCGGGCAAGTACCGGACGCTCATCGCGCGCGCTGA 848
QY 2034 GAATGGAACCTGTATGCTTGGGACCGCGGCTCGGCTGCTTACGCCAGTGG 2093
    || || || || || || || || || || || || || || || || || || || || ||
DB 849 CAGCAGCAACGACGCGGAGCTTACAGCAGC--GTGCGCGGAGCTGAGCTCATGSC 905
QY 2094 AACAGCGCGGCTCTCTGAGGACATCCAGGCTCCAGGGCTTCGCTCCATGACAG 2153
    || || || || || || || || || || || || || || || || || || || || ||
DB 906 CCGGCGCTGACATCAAGACACCTCCCGGGCAACAGTACGGGCTTACAGCGGAC 965
QY 2154 CACCAAGTCCGCGGACGAGCAGCAGCCGCGGCAAGTCTAGTGGGACCAAC 2208
    || || || || || || || || || || || || || || || || || || || || ||
DB 966 CAGCATGGCCAGCCGACGTCGCGGCGCGCTGCACATCTCAGCAAGCAGC 1020

RESULT 11
US-09-408-647A-1
: Sequence 1, Application US/09408647A
: Patent No. 6399858
: GENERAL INFORMATION:
: APPLICANT: Kobayashi, Donald
: TITLE OF INVENTION: Chitinase Gene from Stenotrophomonas
: FILE REFERENCE: maltophilia
: CURRENT APPLICATION NUMBER: US/09/408,647A
: PRIOR FILING DATE: 1998-08-26
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 2810
: TYPE: DNA
: ORGANISM: Stenotrophomonas maltophilia
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (306)...(2405)
: OTHER INFORMATION: Open reading frame 1 (ORF1)
: NAME/KEY: s19-peptide
: LOCATION: (306)...(428)
: NAME/KEY: terminator
: LOCATION: (2457)...(2481)
: NAME/KEY: RBS
: LOCATION: (295)...(298)
: NAME/KEY: CDS
: LOCATION: (314)...(2162)
```

```
: OTHER INFORMATION: Open reading frame 2 (ORF11)
: PUBLICATION INFORMATION:
: DATABASE ACCESSION NUMBER: Genbank No. 6399858 AF014950
: DATABASE ENTRY DATE: 1997-09-23
US-09-408-647A-1
```

```
Query Match 2.4%; Score 60.2; DB 4; Length 2810;
Best Local Similarity 44.7%; Pred. No. 0.00011;
Matches 276; Conservative 0; Mismatches 338; Indels 3; Gaps 1;
```

```
QY 1737 CCTGCCCCGGGCGCGCTATCGCTCGGACAAAGAACCAACAGGCTTTCTAGCGCG 1796
    || || || || || || || || || || || || || || || || || || || || ||
DB 635 CCCCACCGCGGGGCGAGTTCACCGCGGACGACATACCGTCAAGGCCACCGCC 694
QY 1797 CTCCGATGACCTTTTACGTAGCAAGACACCGGACACGTTTACCGCGGGGCCAA 1856
    || || || || || || || || || || || || || || || || || || || || ||
DB 695 CGACACGACGCGAGTGTACGAGGTGAGTGTCTTCCGTGGCGGTACCTCGTGGCAT 754
QY 1857 GCTGGGACGCGGACGAGGACGATCCGGATATTCGCTCCACCGGACCGCGGACGATT 1916
    || || || || || || || || || || || || || || || || || || || || ||
DB 755 CGACACGACGCGCGCTACAGGTGACCTGCGCAATGCAATCGCGCGGACGACACCTT 814
QY 1917 GTATGCTTGACCGACGCTGCGATATTCGCTTCACAGACTGGGACGACCTTTGGCCA 1976
    || || || || || || || || || || || || || || || || || || || || ||
DB 815 CAAAGGCGTGGCCACCGACCAACACCGGCTCACTTCTCGCGACGCTCAGCGTGAC 874
QY 1977 AGTCTCCACCGGCTGACCAACACCTACAGATTCGCGGCTGAGGCTGAGGCTGCA 2036
    || || || || || || || || || || || || || || || || || || || || ||
DB 875 CGTACCGGCTCCAGCAACGACACACCGCGGACGCTGCGCGGCGGCTGCGCTACAC 934
QY 2037 CTGGAACCTGTATGCTTGGGACCGCGGCTGACGAGGCTCGCTTACGCGAGTGGAGA 2096
    || || || || || || || || || || || || || || || || || || || || ||
DB 935 CTCCAAACCGGACACGACGCTGCTGAGCGCGGCGGACACGACAGCGGGG 994
QY 2097 CAGCGGCGCTCTCGGACGACATCCAGGCTCCAGGCG--TTGCGCTCATGACAG 2153
    || || || || || || || || || || || || || || || || || || || || ||
DB 995 CAGTGGCTGGGCGGCTTACGCTACCGCAACGCGACCTGTGCTGCTCACCGCGCC 1054
    || || || || || || || || || || || || || || || || || || || || ||
QY 2154 CACCAAGTCCGCGGACGAGGACGCGCGGCGCAAGTCTAGTGGGACCAACGCGCG 2213
    || || || || || || || || || || || || || || || || || || || || ||
DB 1055 CACCCAGTACACCGACGCGGCGGCTCACCGGACGACGCGCTTACACTTACCGTAC 1114
    || || || || || || || || || || || || || || || || || || || || ||
QY 2214 GGGCGCTTTTACGCTCAGGAAACGCTCGCGGCGGACGAGCGGAGCTTCTGCTGAC 2273
    || || || || || || || || || || || || || || || || || || || || ||
DB 1115 CCGCGACACGCGCGGCAATGCTCGGCGGACGAGCGGCTCGATCAGCTACACCGCGCG 1174
    || || || || || || || || || || || || || || || || || || || || ||
QY 2274 CAAAGACAGACGACGATGCTTTCGCGCAGCTTGACGACACGCTTACGTCAGCT 2333
    || || || || || || || || || || || || || || || || || || || || ||
DB 1175 CGGCGGCGGGGCGGCGGACCAAGCGGCTGATCGCTACTTACCCAGTGGGCGATCTA 1234
    || || || || || || || || || || || || || || || || || || || || ||
QY 2334 TGTATCCAGACCGCGG 2350
    || || || || || || || || || || || || || || || || || || || || ||
DB 1235 CGGCGCACTACCGGG 1251
    || || || || || || || || || || || || || || || || || || || || ||

RESULT 12
US-09-103-840A-1/c
: Sequence 1, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: PRIOR FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 4411529
```


CLASSIFICATION: 510
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4684.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2249 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-814-052-19

Query Match 2.3%; Score 58.4; DB 3; Length 2249;
Best Local Similarity 48.1%; Pred. No. 0.00027;
Matches 201; Conservative 0; Mismatches 211; Indels 6; Gaps 1;

QY 2104 GCCTCTGAGCAGGACATCCAGGAGCTCCAGGGCTTGCGCTCCATCGACAGCACCAGAGTC 2163
DB 1814 GCCTTCGGGGCCGACACCCGGGCGCTGCTGTTCACATGCCCTGGCAGCTC 1873
QY 2164 GCCGGCAGCGGCGACGCGCGGCAAGTCTACGTGGGACCAACGCGCGGCGCTTT 2223
DB 1874 TCGGGCGGCTGGGCTGCTACCTCGAGCGCGCCGACGACCTGGCGGCGGCTCTCG 1933
QY 2224 TACGCTCAGGGAACCTCGCGCGGCGGCGGCGGACCTTCTCTGTCGACCAAGCAGAGC 2283
DB 1934 GACGGCGACCGCGACGACCTCGACCGCTCTGGCGGCGGCGGCGGCTACCTGCTACC 1993
QY 2284 AGCAGCAGTACTCTCTCCGCAAGCTGAGCAGCAGCAGCTGAGGCTGTATCCAGC 2343
DB 1994 AACCCCTACCCCAAGTCCGACCCCTCCAGCAGCAGCAGCTTCGCTCAACCGCTACC 2053
QY 2344 ACCCGGGCTTCGACGGTCTGTCGAGGACCACTGCGCGCGCGCTCCACAGGGGTCA 2403
DB 2054 AGCAGCAGCAGCAGCTGACCTGACCTGAGCGCGCGCAGCTCAGCTACGACTCCG 2113
QY 2404 GGGGTGGCGGGTCATTTAT-----GCTCAGTGGGAGGATTTGGTGGACGGGGCGGAGC 2457
DB 2114 AGCGGCTGACTGCTGAGAGTGGGCTCAGTGGGCGGCAATGGCTGAGAGGGGCTGCACC 2173
QY 2458 CAGTGTGTGGCGCCGCTATGCTGCGCAGAGCAGATGATTATTACTACAGGTGTGT 2515
DB 2174 ACCTGGGTGGCGGACGACTTGCAGAGATTAACTGACTGTACCATCAGTGCCTGT 2231

Search completed: May 4, 2003, 17:26:51
Job time : 22638.3 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 03:05:23 ; Search time 182.021 Seconds
(without alignments)
16339.507 Million cell updates/sec

Title: US-10-026-994-4
Perfect score: 2517
Sequence: 1 atgaagctctcagctcgtcgtgta 2517

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 746054 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121.8	4.8	3668	9	US-09-927-827-21
2	66.8	2.7	1107	10	US-09-748-033-6
3	66.8	2.7	2712	10	US-09-748-033-4
4	64.2	2.6	1635	10	US-09-864-761-20241
5	64.2	2.6	1973	10	US-09-864-761-3471
6	62	2.5	1806	10	US-09-815-242-4015
7	59.2	2.4	1075	10	US-09-864-761-19241
8	59.2	2.4	1403	10	US-09-864-761-2513
9	58.6	2.3	753	10	US-09-887-576-822
10	58.4	2.3	1185	9	US-09-887-576-784
11	58	2.3	1173	9	US-09-894-844-60
12	56	2.2	1248	9	US-09-860-846-7
13	56	2.2	1248	9	US-09-988-384B-7
14	56	2.2	1248	10	US-09-861-289-7
15	56	2.2	12441	9	US-09-988-384B-3
16	56	2.2	13613	9	US-09-860-846-3
17	56	2.2	13613	10	US-09-861-289-3
18	55.8	2.2	68750	9	US-10-014-717-1
19	55	2.2	1896	9	US-10-124-880-15

20	55	2.2	3300	10	US-09-379-931-6	Sequence 6, Appl1
C 21	55	2.2	88421	9	US-09-976-059-1	Sequence 1, Appl1
C 22	54	2.1	520	9	US-10-184-644-332	Sequence 332, App
C 23	54	2.1	520	9	US-10-184-634-332	Sequence 332, App
24	53.4	2.1	7332	10	US-09-844-849-1	Sequence 1, Appl1
25	53.4	2.1	7673	10	US-09-815-925-1	Sequence 1, Appl1
26	52.6	2.1	747	10	US-09-887-576-781	Sequence 781, App
27	52	2.1	88421	9	US-09-976-059-1	Sequence 1, Appl1
C 28	51.8	2.1	1665	10	US-09-815-242-7935	Sequence 7935, Ap
29	51.8	2.1	2274	10	US-09-921-771-5	Sequence 9, Appl1
30	51.6	2.1	1458	9	US-09-860-846-9	Sequence 9, Appl1
31	51.6	2.1	1458	9	US-09-988-384B-9	Sequence 9, Appl1
32	51.6	2.1	1458	10	US-09-861-289-9	Sequence 9, Appl1
33	51.4	2.0	4139	10	US-09-964-824A-105	Sequence 105, App
34	51.4	2.0	4139	10	US-09-964-824A-578	Sequence 578, App
35	51.4	2.0	4139	10	US-09-864-864-334	Sequence 334, App
36	51.4	2.0	4139	10	US-09-880-107-2121	Sequence 2121, Ap
C 37	51	2.0	4341	9	US-10-166-087-45	Sequence 45, Appl
C 38	51	2.0	32539	9	US-10-166-087-1	Sequence 1, Appl1
C 39	50.6	2.0	446	10	US-09-864-761-20699	Sequence 20699, A
40	50.6	2.0	1482	10	US-09-815-242-7857	Sequence 7857, Ap
41	50.4	2.0	390	10	US-09-790-399-7	Sequence 7, Appl1
42	50	2.0	684	10	US-09-927-664-101	Sequence 101, App
43	50	2.0	987	10	US-09-754-014-3	Sequence 1, Appl1
44	50	2.0	6491	10	US-09-997-664-1	Sequence 1, Appl1
C 45	49.6	2.0	4826	10	US-09-772-304A-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-927-827-21
: Sequence 21, Application US/09927827
: Publication No. US200303036176A1
: GENERAL INFORMATION:
: APPLICANT: Bower, Stanley G.
: APPLICANT: Ramseier, Thomas M.
: TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
: FILE REFERENCE: 38-10(15824)B
: CURRENT APPLICATION NUMBER: US/09/927,827
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: US 60/279,493
: PRIOR FILING DATE: 2001-03-28
: NUMBER OF SEQ ID NOS: 69
: SEQ ID NO 21
: LENGTH: 3668
: TYPE: DNA
: ORGANISM: Xanthomonas campestris
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1001)..(2668)
US-09-927-827-21

Query Match 4.8%; Score 121.8; DB 9; Length 3668;
Best Local Similarity 50.8%; Pred. No. 9.4e+25;
Matches 451; Conservative 0; Mismatches 417; Indels 19; Gaps 6;
QY 88 GCGCGCGCGCTTGTCTCCCGCATCATCTTCCATCCCAAGCAAAAGCGCTACATAT 147
DB 1256 GCGCGTGGCGCTTGTCTCCCGCATCATCTTCCCGCATCCCGCAAGCGCTGCGCAT 1315
QY 148 GCACGACAGATATATGGGCGCGCTGACCGCTCAACGCGCGACGACTCAATGACCGCTG 207
DB 1316 GCGCGACCGATGTGGCGCGCTGACCGCTGACCGCTGACCGCGACGCGACGACCGCG 1375
QY 208 ACGATGAGATTCGATATATGCGCGCTGACCAACATGGGCGCATGACGCTTGCCTT 267
DB 1376 CTGACCGACTGTGGCGCGCTGACGACTGGAACATGATGGCGATGACGACATTCGCCCTC 1435
QY 268 GATCCGACGACGATCAAAAGGTATGATCCGCGATGATGATGATGATGATGATGATGAT 327
DB 1436 GACCCCGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATG 1493

OY	328	CCGAGTAATGAGACCATATTCCTCTGTCAGACCCGGGGGCGCAACGTGTCTTACACCAAC	387
Db	1494	-CCGGCACTGCGCGAGTGGTGGCTGGTTCAACCGCGGCGGACGTCGTACGGCTCCAC	15523
OY	388	TTGGCCCTTCAAAAGTCGGGGGTAACTATGCCAGACCGGAGCCGGAGAGCGTGTGGCTGC	447
Db	1553	CTGGCGTTTAAAGTGGGTGTAAACAGCTGTGGCGCGCCAAATGGGAGAGGGGTGGGGTG	16121
OY	448	GATCCGGCCAACTCCAACTATCTACTTTGGTGTCTGTCAAGAAAGGCGCTCTGGAAAG	507
Db	1613	GACCCGACGATGGCCCGCGTGTCTGTGGCTGTGGCGG---GATGGCGGCGCTGTGGCGT	16698
OY	508	TCTACGACGCGCGGCTGACCTTTTCCAAAGTCTCTGTCACGAGCACTGGAGCTAC	567
Db	1670	AGGCACGATGCGGGCGGCGCACTGGCGCAAGTGGCTCTTTTCCGAGCGCGCGCTGGCC	17229
OY	568	ATCCGACACCCGAGTAGTTCCAACGGCTACACAGCGCAAGCAAGACTCATGTGGGTT	627
Db	1730	GGTGCCACCGCGGCGCAATCATGTGGGGCGCGA---GCAGCGCGTGGGAGATCGCTTTGTCT	17868
OY	628	ACGTTACACTCAACCCACACACAGACCGGGGGAGCCAGCTCTGTATCTTTGTGGACG	687
Db	1787	GTGTTCCAGCGACGCACTGGCAACAGCGCTGCGCCAAACCGCGCATCTACGTGGCGGTG	18464
OY	688	GCTGATAACTCACTGCTTACAGTCTATGTGTAGCAGCAATCGCGGCTCCACGTGTGAGTCT	747
Db	1847	TC---CACCGAAGACACGACGACCTGTATGTGTGTCGAAGATGTCCGGCGGAGTTGGGACCG	19033
OY	748	GTACCGGGGCGACGAGGAAATACTTTCTCTACACAGGCCAAACTGCAGCCAGACAGAG	807
Db	1904	GTGGCCGGGCGAACACCGCGGCTGGCGCC---GAGCCCAATGGCGGCGGCGACGATG	19586
OY	808	G--CCTTGTATCTGACCTATTTCCGATGGGACAGAGGCGCTATATGGAACACTTGGCTAG	865
Db	1959	GGCACTGATATCTGACTATATGGGAGACAGCGCGCGGCACTGATGGCGGGGAGGCT	20188
OY	866	TGTGAGGTACGATCTGCAGGGGGAACCTTGGAAAGACATCACCCCTGTCTTGGATAG	925
Db	2019	TGTGGAATTCACGCGCGGACAGGGGGCGGTGTAGATCAGCCGATTTCCGACAGCAG	20788
OY	926	ATCTATATCTTGGCTTGGCGGCGCTTGGCGCTGCATTTGGAAAGCA	972
Db	2079	CCAAATGGGAGTGAATTCGCTGGGGGTGCGGTGGCGGTGATCCGCA	2125
RESULT 2			
US-09-748-033-6			
: Sequence 6, Application US/09748033			
: Patent No. US20020069431A1			
: GENERAL INFORMATION:			
: APPLICANT: Broadway, Roxanne M.			
: APPLICANT: Gongora, Carmenza E.			
: TITLE OF INVENTION: EFFECT OF ENDOCHITTINASE AND CHITOBIOSIDASE AND THEIR			
: FILE REFERENCE: 19603/3091			
: CURRENT APPLICATION NUMBER: US/09/748,033			
: PRIOR FILING DATE: 2000-12-22			
: PRIORITY APPLICATION NUMBER: 60/172,003			
: NUMBER OF SEQ ID NOS: 8			
: SOFTWARE: PatentIn Ver. 2.1			
: SEQ ID NO 6			
: LENGTH: 1107			
: TYPE: DNA			
: ORGANISM: Streptomyces albidoflavus			
US-09-748-033-6			

Query Match	2.7%	Score 66.8;	DB 10;	Length 1107;
Best Local Similarity	44.5%	Pred. NO.3.8e-09;		
Matches 405; Conservative	0;	Mismatches 492;	Indels 13;	Gaps 3

1475 CCACATGGGCCACCTTCGACGACCGTCGACTACGCCGGGACATCGGTCAAGACCGTCGCC 1534

Db	131	CCATCGGTGACAGCTTTGGCCGCTTACGACAAAGCGTACACCCGCCCGCATGTCGTCACG	190
Qy	1535	GCCTGGGCAACACCCGCCGACAGCAACAGATGGCCATCTGTCGACGCGCGCGGACGT	1594
Db	191	GCCTGCGGACACTGGGACCCAGCCGCTGGCGCGCAATTCAACCAAGCTCGCAAGTCA	250
Qy	1595	GGACATCTGACTACGGCGGCCGACACGTCATGAACGCGCGGCAAGGTGGCTATTTCGCCG	1654
Db	251	AGGCCAAGTATACCCGACATCAATGAAGTCTCTGGTCTCTGGCGGGTCGAGCTGTGGTGGCG	310
Qy	1655	ACGGCGACAGCATCTGTGGTTCGACGCGCCGTCGCGCGCTGCAAGCGTCGCGAGTTCCAG	1714
Db	311	GCTTACCGAGACGCGGTGTAACACCCCGCCGCTTCCGCAAGTCTGCGACAGCATGTGCG	370
Qy	1715	GCAGCTTGGCTCCGTCTGACAGCTGCCCGCGGGCGCGCTCATGCTTCGCGACAAGA	1774
Db	371	AGGACCCGCGCTGGGCGGACGCTTTCGACGGGCTACGACTCGACATGAGATACCCGAACG	430
Qy	1775	CCAAAGCGCTTCTTACGCGGGCGCGGATGCACTTTAGCTCAGCAAGGACACCGGCA	1834
Db	431	CTGTGGGCTCATGTTGGACAGCTCCGCGTCCGCGCGCTGTAAGAACATGGTCAAGCGCA	490
Qy	1835	GCAGCTTTCACGCGCGGGGCCCAAGCTGGGACAGCGCAAGGACGATTCGCGGATATCTGCTGC	1894
Db	491	TGGCGCCCATTTGGGACCGGACTGTACACCGCCCATCAACCCCGACCGCACTGCG	550
Qy	1895	ACCCGACACCAGCGCGGACGTTGATGTCTCGACGAGTGGCATTTCCGCTTCACAG	1954
Db	551	GGCGCAAGCTGACGACGCGCGGACTAGC--CGGGCGCGCCCAAGTACATTTCATCTGTACA	607
Qy	1955	ACTGGGACAGCACTTTGGCCAGTCTCCACGCGCTGACCAACACTACAGATCGCC	2014
Db	608	ACGTGATGACGTAGCACTTTTGGCGGCTCTGGAGCAAGACCGGCCGACCGGCCACT	667
Qy	2015	TGGGTGTGGGCTCAGGCTCGAAGTGAACCTGTATGCTCTTGGGACACGCGCCCGTCAAGGG	2074
Db	668	CGGCGCTGAAATCTCTACAGCGGCAATCCCAAGGCGCAATTCACATCGGCGCGGCAATCG	727
Qy	2075	CTCGGCTCTACGCGAGTGGAGACAGGGGGCGCTCTCGAGCAACATMCCAGGGCTCTCCAG	2134
Db	728	CCAAAGCTCAAGGCGAAGGGGTGCCCGGACGAAAGTCTCTGTGAGCATCGGCTTCTAEG	787
Qy	2135	GCCTTCGCTTCATGACAGCAGCAAGGTGCGCGGACGCGACGACCGCGGGCAAG---	2191
Db	788	GCGGGGGGTGAGACGGGTCTACCCAGAGGCGCCCGGGGCGCACCGCACCGCGCGGCA	847
Qy	2192	-----TCTAGTGTGGGACCAAGCGCGGGGGCTCTTTTACGCTCAAGGAACCGTGGCG	2245
Db	848	CCGCGACCTTACGAGGCGGGATCGAGGACTACAAAGTCTCTCAAGAACACTGTCCCGGCA	907
Qy	2246	GCGGACAGGGGGGACTTCTCGTCGACCAAGGACGACGACGATACCTTTCGCGCA	2305
Db	908	CCGGACACCGTGGGCGGACCGCGGTAGCGCCAAAGTGGCGGACAGCACTGTGTGAGCTACGA	967
Qy	2306	GCTCGAGACCAACGCTGAGGTGCA-GCGTTGATCCACGACCGGGGCTTTCGACGCTGACT	2364
Db	968	CCCCGGCGACCATCAAGAACCAAGATGACCTGGGCGCAAGGACCGAGGGCTCTGGCGGCGCT	1027
Qy	2365	TCTGTGAGGA 2374	
Db	1028	TCTTCTGGGA 1037	

RESULT 3
 US-09-748-033-4
 ; Sequence 4, Application US/09748033
 ; Patent No. US2002006931A1
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: Broadway, Roxanne M.
 ;
 ; TITLE OF INVENTION: EFFECT OF ENDOCHITTINASE AND CHITOBIOSIDASE AND THEIR
 ; TITLE OF INVENTION: ENCODING GENES ON PLANT GROWTH AND DEVELOPMENT

FILE REFERENCE: 19603/3091
CURRENT APPLICATION NUMBER: US/09/748,033
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/172,003
PRIOR FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patent Ver. 2.1
SEQ ID NO: 4
LENGTH: 2712
TYPE: DNA
ORGANISM: Streptomyces albidoflavus
US-09-748-033-4

Query Match 2.7%; Score 66.8; DB 10; Length 2712;
Best Local Similarity 44.5%; Pred. No. 4,9e-09;
Matches 405; Conservative 0; Mismatches 492; Indels 13; Gaps 3;

1475 CCACATGGGCCACCTGACGAGCTGACTACCGGGGAACCTGCTCAAGACGCTGCTC 1534
1698 CCATGGGTGACAGCTTCGCGCGCTACGACAAAGCGGTACACCGCGCGAGTGCGTGACG 1757
1535 GCGTGGGCAACACCGCGCGAGCAACAGTGGCCATCTCGTCCGACGGCGCGGACGT 1594
1758 GCGTGGCGGACACCTGAGGACAGCGCTGCGGCACTTCAACAGCTCCGCAAGCTCA 1817
1595 GGAGCATGACTACGCGGCGGACAGCTCCATGACGCGGACAGGTCGCTATTCGGCGG 1654
1818 AGGCCAAGTACCGGCAATCAAGTCTCTGCTCTTCGCGCGGTGAGCTGTCGGGG 1877
1655 ACGGCAACAGTCTCTGCTGAGCGCTCTCGCGGCTGACAGCGCTGCAAGTTCAGG 1714
1878 GCTTACCCACCCGCTGGAAGAACCGCGCGCTTCGCAAGTCTGCGCACAGCTGCTG 1937
1715 GGAGCTTTCCTCGCTGACGCTGCGCGGGGCGCGCTCATCGCTGGGACAAAGA 1774
1938 AGGACCGCGCTGGCGGACGCTTTCGACGAGCATGACCTCACTGGGAGTACCGGAACG 1997
1775 CCACAGCGCTTCTTACCGCGCGCTCGATGACCTTTACGTCACAGACAGACCGGCA 1834
1998 CCGTGGGCGCTGAGTGGGACAGCTCGGTCGCGCGCGCTGGAAGACATGTCAGAGCA 2057
1835 GCACTTACGCGGCGGCGCAAGCTGAGGACGAGGAGCATCGCGGATTCGCTGCTC 1894
2058 TCGCGGCCAGTTCGAGCAGCTGCTGACCGCGCCATCACCGCGGACGAGCTCGG 2117
1895 ACCCGACACCGCGGCGAGTGTGATGTCAGCAGAGCTCGGCAATTCGCTCCACAG 1954
2118 GCGGCAAGCTCGACGCGCGCACTACG--CGGGCGCGCGCAGTACTGACTGCTACA 2174
1955 ACTGGGCGACGACTTTGGGCAAGTCTCCACCGCGCTGACCAACACTACGAGATGCC 2014
2175 AGGTATGACGATACGACTTCTTCGGGCGCTGGGACAAAGACCGCGGACCGCGCACT 2234
2015 TGGGTGTGGGCTCAGGCTGGAACCTGTATGCTTCGGGACCGCGCGCTCAAGGG 2074
2235 CGGCGCTGAACTCTCAACAGCGGATCCCAAGCGCACTTCACTGGCGCGCATG 2294
2075 CTCGCTTACCGCAAGTGAAGACAGCGGCGCTCTGAGACGAGATCCAGGGCTCCAGG 2134
2295 CCAAGGTCAAGCGAAGGCGGTCCGCGGAGCAAGCTCTGCTCGGCAATGCTTCTAGG 2354
2135 GCTTGGGCTCATGACAGCAACAGGTGCGGACGAGACCGCGGCGAAG--- 2191
2355 GCGGCGGTGGAGCGCGGTACACCAAGAGCGCCCGGGCGGACCGCGCGGCGCA 2414
2192 -----TTAGCTGGGCAACAGGCGCGGCGCTTTTACGCTCAAGGACGCTCGGCG 2245
2415 CCGGCACTACGAGCGGCGGAGTCAAGAGTCTCAAGAGTCTCAAGAGTCTCGCGCA 2474
2246 GGGGCAAGGCGGAGTCTCTGTCAGCAAGAGAGAGCAAGTACTCTTCGGCA 2305
2475 CCGGCAAGTCTCGGCGGCGGACCGGTCAGCAAGTTCGCGGAGCAAGTGTGAGTACGACA 2534

2306 GCTGAGCACACGCTGAGTGA-GCGTTGTATCCACGACCGCGGCTTCGAGGTGACT 2364
2535 CCGCGGCGACATCAAGACCAAGATGACTGGGCCAAGGACCAAGGCGCTCGCGGCGCT 2594
2365 TCGTGAAGA 2374
2595 TCTTGGGA 2604

RESULT 4

US-09-864-761-20241
Sequence 20241, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn. Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aegmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 20241
LENGTH: 1635
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006547.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 8.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-20241

Query Match 2.6% Score 64.2; DB 10; Length 1635;
Best Local Similarity 42.3%; Pred. No. 2.3e-08;
Matches 415; Conservative 0; Mismatches 563; Indels 3; Gaps 1;

1419 CACCTTTGCCAGAGAAACACCTCGGAGATCGCCGACACCGTCTGGGCAACGCCAC 1478
1419 CACCTTTGCCAGAGAAACACCTCGGAGATCGCCGACACCGTCTGGGCAACGCCAC 1478
379 CACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCAT 438
1479 ATGGGCACTCGACAGAGCTGCACTACCGGGAACTCGGTAAGAGCTGCTCCGCT 1538
1479 ATGGGCACTCGACAGAGCTGCACTACCGGGAACTCGGTAAGAGCTGCTCCGCT 1538
439 CACTACTACACACACACACACACACACACACACACACACACACACACACACAT 498
1539 CGGCAACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1598
499 CACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCAT 558
1599 CACGATCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1658
559 CACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCAT 618
1659 CGACAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1717
619 CACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCAT 678
1718 -GCTTGGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1775
679 TGGCACCACACACACACACACACACACACACACACACACACACACACACACAT 738
1776 CAACAGCTCTTCTAGCGCGGCTCGGATGACCTTTTACGAGAGAGACCGGACG 1835
739 CACTACACACACACACACACACACACACACACACACACACACACACACACAT 798
1836 CAGCTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1895
799 CACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCAT 858
1896 CCGGACACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1955
859 CACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCAT 918
1956 CTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2015
919 CACTACTACACACACACACACACACACACACACACACACACACACACACACAT 978
2016 GGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2075
979 CACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCAT 1038
2076 TGGCCTTCTAGCGCGGCTCGGATGACCTTTTACGAGAGAGACCGGCGGCGG 2135
1039 TACACACACACACACACACACACACACACACACACACACACACACACACAT 1098
2136 CTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2195
1099 CACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCAT 1158
2196 CTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2255
1159 CACTACTACACACACACACACACACACACACACACACACACACACACACAT 1218
2256 CCGGACTTCTGCTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2315
1219 CACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCAT 1278
2316 CAGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2375
1279 CACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCATTAACCAT 1338

2376 CAGCTGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2396
1339 CACTACTTCCACACACACACACACACACACACACACACACACACACACACAC 1359

US-09-864-761-3471
Sequence 3471, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmlca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO. 3471
LENGTH: 1973
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006547.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-3471

```
Query Match      2.6%; Score 64.2; DB 10; Length 1973;
Best Local Similarity 42.3%; Pred. No. 2.5e-08;
Matches 415; Conservative 0; Mismatches 563; Indels 3; Gaps 1;

QY 1419 CACCTTTGCGAGAGAAACGCTCGGAGATCGCCGAGACGCTGTGGGCAAGGCCAC 1478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 533 CACCATACCATCATACATACCATACCATACCATACCATACCATACCATACCATAC 592
QY 1479 ATGGGCGACCTGACAGAGCTGCATACCGCGGAAGTGGTCAAGAGGCTCTCCGCGT 1338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 593 CATCACTACCATACCATACCATACCATACCATACCATACCATACCATACCATAC 652
QY 1539 CGGCAACACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 653 CACCATCACTACCATACCATACCATACCATACCATACCATACCATACCATACCAT 712
QY 1599 CATGCACTACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 713 CACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATAC 772
QY 1659 CGACAGATCTCTGCTGCAAGCGCTGTCGCGGCTGCAAGCGCTGCAAGGAGCA- 1717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 773 CACCATCACTACCTGCGACACCATACCATACCATACCATACCATACCATACCATAC 832
QY 1718 -GCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1775
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 833 TGGCACCACACACACACACACACACACACACACACACACACACACACACACACAT 892
QY 1776 CACAGAGCTCTTCTACGCGGCTCCGAGTGCAGCTTTTACGTGAGAGAGACCGGAG 1835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 893 CACTACACACACACACACACACACACACACACACACACACACACACACACACAC 952
QY 1836 CAGCTTACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1895
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 953 CACCATCACTACCATACCATACCATACCATACCATACCATACCATACCATACCAT 1012
QY 1896 CCGGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1955
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1013 CACCATCACTACCATACCATACCATACCATACCATACCATACCATACCATACCAT 1072
QY 1956 CTCGGGCGAGACCTTTGGCAAGTCTCAACCGGCTGACCAACACCTTACAGATCCGCT 2015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1073 CATCACTACCATACCATACCATACCATACCATACCATACCATACCATACCATAC 1132
QY 2016 GGGTGTGGGCTAGGCTGCAATGGAAGCTGTATGCTTGGGACCGGCGGCTGAGGGG 2075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1133 CACCATCACTACCATACCATACCATACCATACCATACCATACCATACCATACCAT 1192
QY 2076 TGGCTCTACGAGTGAAGAGAGCGGCGGCTGAGAGGAGATCAAGGGCTCCAGGG 2135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1193 TACACACACACACACACACACACACACACACACACACACACACACACACACAT 1252
QY 2136 CTTCGGCTCTACGAGAGACCAAGGTGCGCGGAGAGCGGAGCGGCGGCGGAGTCTA 2195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1253 CACCATCACTACCATACCATACCATACCATACCATACCATACCATACCATACCAT 1312
QY 2196 CCGGGCGACCAAGCGGCGGCGGCTTTTACGTCAAGGAGACCGTGGGCGGCGGAG 2255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1313 CATCACTACCATACCATACCATACCATACCATACCATACCATACCATACCATAC 1372
QY 2256 CGGAGCTTCTGCTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1373 CACCATCACTACCATACCATACCATACCATACCATACCATACCATACCATACCAT 1432
QY 2316 CAGCTGAGGTGAGGCTGTATCAAGAGCGGCGGCTGAGAGGAGTCTGCTGAGGAG 2375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1433 CACCATCACTACCATACCATACCATACCATACCATACCATACCATACCATACCAT 1492
QY 2376 CAGCTGAGGCGGCGGCTGAGG 2396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
DB 1493 CATCACTTCCACCCACCCAC 1513

RESULT 6
US-09-815-242-4015
; Sequence 4015, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4015
; LENGTH: 1806
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4015

Query Match      2.5%; Score 62; DB 10; Length 1806;
Best Local Similarity 44.8%; Pred. No. 1e-07;
Matches 239; Conservative 0; Mismatches 295; Indels 0; Gaps 0;

QY 1689 CGGCTGTCAGCGCTGCGAGTTCAGGGGAGGCTTGCCTCCGCTCGAGGCTGCCGCGG 1748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 339 CAGCGAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 398
QY 1749 CGCGGCTATCGGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 399 CGCATGAAACCTCTGATAGCCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 458
QY 1809 CTTTTACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1868
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 459 GCGTGTCTCGGCGAGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 518
QY 1869 AGGAGAGATCGGGATATGCTGCTACCCGAGCAGCGGCGGCGGCGGCGGCGGCGG 1928
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 519 CCGGAGCATGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 578
QY 1929 CGAGCGGCGAGATTCGCTGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 579 CACCAAGATCTTATCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 638
QY 1989 CCGTACCAACACCTACAGATGCGCTGTGGGTGCGCTCAGGCTGCAAGTGAACCTGA 2048
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 639 CCGTGGCGCGCTGCGGAGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 698
QY 2049 TGCCTTGGGCGGCGGCGGCTGCAAGGGGCTGCTTACGCGCAAGTGAAGAGAGCGGCG 2108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 2513
LENGTH: 1403
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL078472.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
US-09-864-761-2513

Query Match 2.4%; Score 59.2; DB 10; Length 1403;
Best Local Similarity 43.0%; Pred. No. 6.1e-07;
Matches 289; Conservative 0; Mismatches 383; Indels 0; Gaps 0;

QY 1725 CTCGCTCTGAGCTGCCCGGGGGCGCTCATCGCTTCGGACAGAGACCAACAGCCT 1784
DB 1122 CACCCACACACATCACCACATCACCACATCACCACATCACCACATCACCACATC 1063
QY 1785 CTTTACCGCGGCTCCGATCGACCTTTTACGTGACAGAGACAGAGAGAGAGAGAG 1844
DB 1062 CACCCACACACATCACCACATCACCACATCACCACATCACCACATCACCACATC 1003
QY 1845 GCGCGGCGCCAGCTGGGCGAGGAGGAGATCGGGATATCGCTGCTACCCGACAC 1904

DB 1002 CACTACACACACACACACACACATCACCACACACACACACATCACCACACACACAC 943
QY 1905 CCGGGGACAGCTTTATGTCGACCGAGAGTGGCATTTCCGCTCCACAGACTGGGAC 1964
DB 942 CACCCACACATCACCACATCACCACACACACATCACCACATCACCACATCACCAC 883
QY 1965 GACCTTTGGCCAAAGTCTCCACCGGCGTGGACCAACAGCTACAGATGCGCTGGTGG 2024
DB 882 CACCCACACACATCACCACATCACCACACACACACACACATCACCACATCACCAC 823
QY 2025 CTCAGGCTCGAAGCTGAACTGTATGCTTCGGACACCGGCGGCTGAGGGGCTGAG 2084
DB 822 CACTACACATCACCACACACACACACATCACCACATCACCACATCACCACATCACC 763
QY 2085 CCGCAGTGGAGACAGCGGCGCTCTGGAGGAGACATCCAGGGCTCCAGAGGCTTGG 2144
DB 762 CACCCAGCACACACATCACCACACACACACACATCACCACATCACCACATCACC 703
QY 2145 CATCGACACACACAGTGGCGGCGGAGGAGGAGCGGCGGCAAGTCTACGTGGGAC 2204
DB 702 CATCACCACACACACATCACCACATCACCACATCACCACACACACATCACCACAT 643
QY 2205 CACGCGCGGGCGCTTTTACGTCAGGGAACCGTCCGCGGCGGACGCGGCGGACTTC 2264
DB 642 CATCACCACACATCACCACACACACATCACCACATCACCACACACACATCACCAC 583
QY 2265 CTCGTCGACACAGACAGACAGACAGATCTCTTCGCGGAGCTGAGACACACCTGAG 2324
DB 582 CACCATCACCACATCACCACACACACATCACCACATCACCACATCACCACATCACC 523
QY 2325 GTCGAGCGTGTATCCAGACCGGCGGCTTCGAGGAGTGTGAGGAGACAGTGGG 2384
DB 522 CACCCACACACATCACCACACACATCACCACATCACCACACACACATCACCACAC 463
QY 2385 CGCGGTCCAC 2396
DB 462 TACCATCACCAC 451

RESULT 9
US-09-887-576-822
Sequence 822, Application US/09887576
Patent No. US20020144047A1
GENERAL INFORMATION:
APPLICANT: Budworth, P.
APPLICANT: Brown, D.
APPLICANT: Chang, H.
APPLICANT: Zhu, T.
APPLICANT: Han, B.
APPLICANT: Wang, X.
APPLICANT: Cooper, Bret
TITLE OF INVENTION: Promoters for regulation of plant expression
FILE REFERENCE: 1360,001051
CURRENT APPLICATION NUMBER: US/09/887,576
CURRENT FILING DATE: 2001-06-25
PRIOR APPLICATION NUMBER: US 60/213,848
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/214,087
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: US 60/258,692
PRIOR FILING DATE: 2000-12-29
NUMBER OF SEQ ID NOS: 875
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 822
LENGTH: 753
TYPE: DNA
ORGANISM: Oryza sativa
US-09-887-576-822

Query Match 2.3%; Score 58.6; DB 10; Length 753;
Best Local Similarity 45.3%; Pred. No. 7.6e-07;
Matches 257; Conservative 0; Mismatches 304; Indels 6; Gaps 1;

Oy	1750	GGCGTCAATCGCTCGGACAAAGAACCAACAGCGCTTTCTACGGCGGCTCGGAAATCGAAC	1809
Db	64	GGCGGCTTGGCGGAGTTCAATCTCCACCTTCATCTTGTCTTGCGCGGCGCAAGGGCTCGCGG	123
Oy	1810	TTTTACGTACGACAAAGACACCGGCGACAGCTTTCACGCGGGGCGCAAGCTGGGACGCGCA	1869
Db	124	ATGGGCTTTGACGAACGTGACGGGGGGGGAGGAGACGACACCGCGCGGGCTGATTCGGCGG	183
Oy	1870	GGGACGATCGGGGATATCGTCTCAACCGGACACCGCGGGGACGTTGATGTCTGACG	1929
Db	184	GGGTTGGCGCCACGCGCTTTCGGCCCTGTTCTGTTGGCGGTTCCTCGCGGCGCAATCTCGCG	243
Oy	1930	GACGTGGGCAATTTCCGCTCCACAGACTGGGGAGACCTTTGGCCAACTCTCCACCGCC	1989
Db	244	GGCCACGTGAACCCCGCGCGCTGAGCTTTCGGCGGCTTCTGTTGGGCGCAACATCACTCTGTTC	303
Oy	1990	CTGACCAACACCTTCACTGATCGGCTTGGGAGTGGGCTCAAGCTCGAATCGAATCTGTAT	2049
Db	304	CGGGGCTCTCTACTGTGATTCGCCAGCTCTCTGGCTTCCACCTGGGCTGTTCTCTCTC	363
Oy	2050	GCGTTTGGGACACGGCGCCGCTGACGAGGGGCTCGGCTTACGCGCAGTGAGAGACAGCGGCGCTCC	2109
Db	364	CGCTTCTTCACCGCGGCGGCTCGCGMACCGGACCTTGGCCCTAACCGGCGTTCGCTGG	423
Oy	2110	TGGACGGAATCCAGGGCTCTCCAGGGCTTTCGGCTCATGTGAAGACCAAGTGGCGGCG	2169
Db	424	GAGGCGCTTGCGTGGTGGAGATGTCATGACTTTCGGCTCTGTCTAACACGCTGTACACGCC	483
Oy	2170	AGCG-----GCAACACCGCGCGGCAAGTCTACGTGGGACCAACAGCGCGGGGCGCTTTT	2233
Db	484	GGCGTTTGACCCCAAGAAAGGGGACCTCTCGGACCATTCGCCCATATGGCTTTCATC	543
Oy	2224	TAGCGCTAGGGAACCGTGGCGGCGGCGGACGAGGGGACTTCTCTGTCAACCAAGACAGC	2283
Db	544	GTCGGCGCGCAACATCTCTGTCGGCGGCGGCTTGACAGCGCGCTTCATGAACCGGCGCTC	603
Oy	2284	AGCAGCACTACTCTTTCGGCGCACTCG	2310
Db	604	TTCCTTGGCGCCGCGCTTCGTCACTGG	630

```

Query Match      2.3%   Score 58.4   DB 10: Length 1185;
US-09-887-576-784
:
: Sequence 784, Application US/09887576
: Patent No. US2002014047A1
: GENERAL INFORMATION:
: APPLICANT: Budworth, P.
: APPLICANT: Brown, D.
: APPLICANT: Chang, H.
: APPLICANT: Zhu, T.
: APPLICANT: Han, B.
: APPLICANT: Wang, X.
: APPLICANT: Cooper, Bret
: TITLE OF INVENTION: Promoters for regulation of plant expression
: FILE REFERENCE: 1360.001us1
: CURRENT APPLICATION NUMBER: US/09/887,576
: CURRENT FILING DATE: 2001-06-25
: PRIOR APPLICATION NUMBER: US 60/213,848
: PRIOR FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 60/214,087
: PRIOR FILING DATE: 2000-06-23
: PRIOR APPLICATION NUMBER: US 60/258,692
: PRIOR FILING DATE: 2000-12-29
: NUMBER OF SEQ ID NOS: 875
: SOFTWARE: fastseq for Windows Version 4.0
: SEQ ID NO 784
: LENGTH: 1185
: TYPE: DNA
: ORGANISM: Oryza sativa

```

Best Local Similarity 43.88; Pred. No. 9,8e-07;
Matches 351; Conservative 0; Mismatches 446; Indels 5; Gaps 2

Oy	1527	CGTGTCCGCGTGGGCAACACCCCGGACGCAACAGTGGGCATCTGTCGACAGGGG	1588
Db	168	CGGGGAATCAACCAACAGGCGACCGCTCACTACGGAAGATGTCGCGACACTTCCG	227
Oy	1587	CGCAGCTGGAGCAATCCACTACCGGCGCACAGTGCATGAACGGCGGACGGTGGCCTA	1646
Db	228	CGGATGGGCTTCTGTGTCGACGACGTCGGCCCTCGACGGCGACCGCTGCMAAGTCTGT	287
Oy	1647	TTTGGCGGAGCGGCGACACGATCTCTTGTCGACCGGCTGTCGGGCTGACAGGCTCGCA	1706
Db	288	CAACATGAGAGCAGCAGTCGCCCGCAATCCGCGACGGGGGTGCAGCGGCACTTACCACAGC	347
Oy	1707	GTTCCAGGGGAGCTTTCGCTTCGCTGTCGAGCCGCCCGGGGGGCGGTCATTCGCTCGA	1766
Db	348	CCCCGAGAGTATCGGCGCCGGCGACCAAGGGCCCCATGTTTCGGCTACGGCACCGAGAC	407
Oy	1767	CAAGAGACCAACAGCGCTCTTCTACGCGCGGCTCCGGATCGACCTTTTACGTCAGCAAGA	1826
Db	408	CCCCGAGCTGATGCTTCCTTACGACCACTCTTCGCGACCAAGCTTCGGCGCGCGCTACCGA	467
Oy	1827	CACCGCAGCAGCTTTCACGCGCGGGGCCAAGCTGGGCGACGCGAGGAGATCCGGGATAT	1886
Db	468	GGTCCGGCAAGACGGGCACTGGCG---CTTGGTCTCAAGCCCGGACGAGACCAAGCTAC	524
Oy	1887	CGGTGTACCCCGGACCAACGCGGGGCAAGTTTATGTCCTGACCGAGCTGCGCATATTCG	1946
Db	525	CGTTTGAGTACCTCAACGACCGCGCGCATGTTCGCCGCTCCGCTCACACCTCTCTAT	584
Oy	1947	CTCCACAGACTCGGGCGACGACTTTTGGCCAAAGTCTTCCACCGGCTTGACCAACCTTACA	2006
Db	585	CTCCACCCAGCAGCAGACGAGACGTCACCAAGCAGCAGATGCGCGCGCACTCAAGAGACA	644
Oy	2007	GATGCGCCTGGGTGTGGGGCTCAGGCTTCGAAGTGAACCTGTATGCTTTCGGCACCGGCC	2066
Db	645	CGTATCAAGCGGCTCTATCCCGACAAAGTACTCGACGAGAAGACATCTTCCACCTCAA	704
Oy	2067	GTCAGGGGCTCGCCTCTACGCCAGTGGAGAACAAGGGCGGCTCTGAGAGGACATCCAGG	2126
Db	705	CCCCTCGGGCGGCTTCTGCTCATTCGGCGGGGCCACAGGCGAGCGCGGCTCACCGGCGAA	764
Oy	2127	CTCCCAAGGCTTGGGCTCATTCGACAGCACCAAGTTCGCGGCGACGGCAGCACCGCGCG	2186
Db	765	GATCATCATCGACACCTCAGCGGAGATGGGGCGGCGACGCGGCGGCGCTTCTCCGCAA	824
Oy	2187	GCAAGTCTACGTGGGACCAAGGGCGGGGCGCTTTTACGCTCAGGGAGACCGTCGGGG	2246
Db	825	GGACCCGACCAAGGTGCACCGGAGCGGCGGCTTACATCGCCAGGCAAGGCGCCACAGAGAT	884
Oy	2247	CGGACACGGGCGGACCTTCTCTGTCGACCAAGCAGAGACAGACAGTACTCTTCGCGCAG	2306
Db	885	CGTGCGCAGGGG--CTCTGCGCGCGCTGCGATGTCAGAGTGTGTAAGCCATGGGCTC	942
Oy	2307	CTCAGACCAACAGCTGAGTCTG	2328
Db	943	CCGAGCGCGCTCTCCGTGTC	964

```

RESULT 11
US-09-894-844-60/c
; Sequence 60, Application US/09894844
; Patent No. US20020176873A1
; GENERAL INFORMATION:
; APPLICANT: Behr, Marcel
; APPLICANT: Small, Peter
; APPLICANT: Schoolnik, Gary
; APPLICANT: Wilson, Michael A.
; TITLE OF INVENTION: Molecular Differences Between Species of
; TITLE OF INVENTION: the M. Tuberculosis Complex
; FILE REFERENCE: STAN102CON
; CURRENT APPLICATION NUMBER: US/09/894,844

```



```
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 7
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-7

Query Match
2.2%; Score 56; DB 9; Length 1248;
Best Local Similarity 43.9%; Pred. No. 4.8e-06;
Matches 338; Conservative 0; Mismatches 425; Indels 7; Gaps 2;

QY 1623 CATGACGGGGGACGGTGGCTTATTTGGCCGACGGGACACAGCTCTGTGTCGACCGC 1682
DB 84 CATGACGGGGGACGGTGGCTTATTTGGCCGACGGGACACAGCTCTGTGTCGACCGC 143
QY 1683 CTGCTCCGCGGCTGACGGCTTCGACAGGAGGAGCTTTGCTTCGTCGACCTGCGC 1742
DB 144 CAACGGGGGGCCGCTCTCCGCGAGTTGAGAGAGCGGCTCGCGGGGCTCGCGGGGTCG 203
QY 1743 CGCGGGCGCGCTATCGCTTCGACAGAGAACCAACAGCGCTTCTTACGCGCGCTCGG 1802
DB 204 GCATGCGCGTGGCACCTGCAACGCGCGCGGCTCCAGCTCTCTCGCGACGCGCGCGG 263
QY 1803 ATGACCTTTTATGTCGACAGAGACACGGGACAGAGCTTACGCGGGGCGCCCAAGCTGG 1862
DB 264 CCTCACCGGGAAGTATGATGCGGTGATGATGATGATGATGATGATGATGATGATGATG 323
QY 1863 CACGCGAGGAGACATCGGGATATCGCTGCTCACCGGACGCGGCGGACGTTGTATGT 1922
DB 324 CTGAGTCGGGCTCACCGCGGCTTTCGCGGACATGACCGCGGACGCGGACGCGGACGCG 383
QY 1923 CTGAGCGAGCTGGGATATTCGCTTCACAGACTCGGGGACGACGCTTTGGCCAAAGTCTC 1982
DB 384 GGACAGAGTGGCGCGCGGCTTCACACCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGG 443
QY 1983 CACGCGCGGACGACGATCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGG 2042
DB 444 GGGCGCGCGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
QY 2043 CTTGATGCTTTCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2102
DB 501 GCTGATGCTTTCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560
QY 2103 CGGCTCTGAGGACATCCAGAGGCTCCAGGCGCTTCGCTTCAGTACGACAGCAAGGT 2162
DB 561 CTTGAGGAGGACGCGAGGCTTTCAGGCTTCAGGCGCGCGCGCGCGCGCGCGCGCGCG 620
QY 2163 CGCGGAGGAGGAGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2222
DB 621 CGCGGCGCGCTTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 680
QY 2223 TTAAGCTCAGGAGACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2282
DB 681 CGGCTTCGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 740
QY 2283 CACGACG---AGTACTCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2338
DB 741 CGCGCGCGCGCATGGGCTTCAGCTTCCTCGAGCGGTTTCCGAGGTCATCACCGGAACG 800
QY 2339 CACAGACCGCGGCTTCGAGCGGTCGTCGAGAGACGACGCTCGCGGCGCGCGCGCGCG 2388
DB 801 CGCGACACGCGCGCTACCGCGAGGACGCTCGCGGAGCTCCCGCGCGCTCC 850
```

```
RESULT 14
US-09-861-289-7
; Sequence 7, Application US/09861289
; Patent No. US20020110897A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Xue, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FASTSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-7

Query Match
2.2%; Score 56; DB 10; Length 1248;
Best Local Similarity 43.9%; Pred. No. 4.8e-06;
Matches 338; Conservative 0; Mismatches 425; Indels 7; Gaps 2;

QY 1623 CATGACGGGGGACGGTGGCTTATTTGGCCGACGGGACACAGCTCTGTGTCGACCGC 1682
DB 84 CATGACGGGGGACGGTGGCTTATTTGGCCGACGGGACACAGCTCTGTGTCGACCGC 143
QY 1683 CTGCTCCGCGGCTGACGGCTTCGACAGGAGGAGCTTTGCTTCGTCGACCTGCGC 1742
DB 144 CAACGGGGGGCCGCTCTCCGCGAGTTGAGAGAGCGGCTCGCGGGGCTCGCGGGGTCG 203
QY 1743 CGCGGGCGCGCTATCGCTTCGACAGAGAACCAACAGCGCTTCTTACGCGCGCTCGG 1802
DB 204 GCATGCGCGTGGCACCTGCAACGCGCGCGGCTTCAGCTCTCTCGCGGACGCGCGCGG 263
QY 1803 ATGACCTTTTATGTCGACAGAGACACGGGACAGAGCTTACGCGGGGCGCCAAAGCTGG 1862
DB 264 CTTACCGGGAAGTATGATGCGGTGATGATGATGATGATGATGATGATGATGATGATG 323
QY 1863 CACGCGAGGAGACATCGGGATATCGCTGCTCACCGGACGCGGCGGACGTTGTATGT 1922
DB 324 CTGAGTCGGGCTCACCGCGGCTTTCGCGGACATGACCGCGGACGCGGACGCGGACGCG 383
QY 1923 CTGAGCGAGCTGGGATATTCGCTTCACAGACTCGGGGACGACGCTTTGGCCAAAGTCTC 1982
DB 384 GGACAGAGTGGCGCGCGGCTTCACACCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGG 443
QY 1983 CACGCGCGGACGACGATCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGG 2042
DB 444 GGGCGCGCGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
QY 2043 CTTGATGCTTTCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2102
DB 501 GCTGATGCTTTCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560
QY 2103 CGGCTCTGAGGACATCCAGAGGCTCCAGGCGCTTCGCTTCAGTACGACAGCAAGGT 2162
DB 561 CTTGAGGAGGACGCGAGGCTTTCAGGCTTCAGGCGCGCGCGCGCGCGCGCGCGCGCG 620
QY 2163 CGCGGAGGAGGAGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2222
DB 621 CGCGGCGCGTTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 680
QY 2223 TTAAGCTCAGGAGACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2282
DB 681 CGGCTTCGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 740
QY 2283 CACGACG---AGTACTCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2338
DB 741 CGCGCGCGCGCATGGGCTTCAGCTTCCTCGAGCGGTTTCCGAGGTCATCACCGGAACG 800
QY 2339 CACAGACCGCGGCTTCGAGCGGTCGTCGAGAGACGACGCTCGCGGCGCGCGCGCGCG 2388
```


Db 801 GCGCAACCAAGCCGCTTACCGGAGACACTCGCGGACCTCCCGGCTCC 850

RESULT 15

US-09-988-384B-3/c
; Sequence 3, Application US/09988384B
; Publication No. US20030073824A1

; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.536US1

; CURRENT APPLICATION NUMBER: US/09/988,384B

; CURRENT FILING DATE: 2001-11-19

; PRIOR APPLICATION NUMBER: PCT/US99/14398

; PRIOR FILING DATE: 1999-06-25

; PRIOR APPLICATION NUMBER: US 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 53

; SEQ ID NO 3

; LENGTH: 12441

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-988-384B-3

Query Match 2.2% Score 56: DB 9: Length 12441:

Best Local Similarity 43.9%: Pred. No. 9.3e-06;

Matches 338: Conservative 0: Mismatches 425: Indels 7: Gaps 2:

QY 1623 CATGACGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 1682
DB 12113 CATGACGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 12054
QY 1683 CTCGTCGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 1742
DB 12053 CACGCGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 11994
QY 1743 GCGGCGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 1802
DB 11993 GATGCGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 11934
QY 1803 ATGACCTTTTACGTACGACGAGACACCGGCGGAGACAGATCTGTGTCACCGC 1862
DB 11933 CCTCACCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 11874
QY 1863 CAGCGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 1922
DB 11873 CTGGATCGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 11814
QY 1923 CTGACCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 1982
DB 11813 GAGCGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 11754
QY 1983 CACGCGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 2042
DB 11753 GGGCGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 11697
QY 2043 CCTGTATGCTTGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 2102
DB 11696 GCTGTATGCTTGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 11637
QY 2103 GCGCTTCGAGGAGATCCAGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 2162
DB 11636 CCTCGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 11577
QY 2163 CGCGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 2222
DB 11576 CGCGGCGGCGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 11517
QY 2223 TTACGCTGAGGAGATCCAGGAGTGGCTATTGCGCGGCGGAGACAGATCTGTGTCACCGC 2282

Db 11516 CGGCTTCGACTGCCCCGCGGAGCGCCCGCGGAGACCAAGCATGAGCGAGGC 11457
QY 2283 CAGCAGC---AGTACCTTCGCGCAGCTCGAGCAGCAGCTGAGGTCGAGCTTGTAT 2338
DB 11456 CGCGGCGGCGGAGTGGCTATTGCGCGGCGGAGTGGCTATTGCGCGGCGGAGCAGC 11397
QY 2339 CCACGACCGGCGGCTTCGAGGTCGTCGAGCAGCAGCTGCGCGGC 2388
DB 11396 GCGCAACCAAGCGGCGGCTTCGAGGTCGTCGAGCAGCAGCTGCGCGGC 11347

Search completed: May 4, 2003, 06:51:25
Job time : 232.021 secs

THIS PAGE BLANK (USP10)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2003, 06:18:59 ; Search time 2307.53 Seconds
(without alignments)
17665.653 Million cell updates/sec

Title: US-10-026-994-4
Perfect score: 2517
Sequence: 1 atgaagctctcagctcgtcgtgtgtga 2517

Scoring table:
IDENTITY: NUC
Gapop 10.0, Gapept 1.0

Searched: 16154066 seqs, 809774376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estbda: *
2: em_esthum: *
3: em_estlin: *
4: em_estmuv: *
5: em_estmuv: *
6: em_estmuv: *
7: em_estmuv: *
8: em_estmuv: *
9: gb_est1: *
10: gb_est2: *
11: gb_est3: *
12: gb_est4: *
13: gb_est5: *
14: gb_est6: *
15: em_estfun: *
16: em_estfun: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rnd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	281	11.2	546	12	BF072664 NCSM3H7T3
2	135.4	5.4	440	14	BO751759 EST632322
3	111.4	4.4	426	14	BO751758 EST632321
4	66.4	2.6	925	17	CNS0091P AL053013 Drosophila
5	64.6	2.6	699	10	BE041072 OF19D05 O
6	62.4	2.5	445	13	BI306418 BI306418 NL_4_J11

7	62.4	2.5	458	13	BI305544
8 <td>60.8<td>2.4<td>925<td>17<td>CNS0091P</td></td></td></td></td>	60.8 <td>2.4<td>925<td>17<td>CNS0091P</td></td></td></td>	2.4 <td>925<td>17<td>CNS0091P</td></td></td>	925 <td>17<td>CNS0091P</td></td>	17 <td>CNS0091P</td>	CNS0091P
9 <td>60.2<td>2.4<td>559<td>14<td>BO463043</td></td></td></td></td>	60.2 <td>2.4<td>559<td>14<td>BO463043</td></td></td></td>	2.4 <td>559<td>14<td>BO463043</td></td></td>	559 <td>14<td>BO463043</td></td>	14 <td>BO463043</td>	BO463043
10 <td>59.5<td>2.3<td>629<td>14<td>BO620053</td></td></td></td></td>	59.5 <td>2.3<td>629<td>14<td>BO620053</td></td></td></td>	2.3 <td>629<td>14<td>BO620053</td></td></td>	629 <td>14<td>BO620053</td></td>	14 <td>BO620053</td>	BO620053
11 <td>56.6<td>2.2<td>596<td>10<td>BE419335</td></td></td></td></td>	56.6 <td>2.2<td>596<td>10<td>BE419335</td></td></td></td>	2.2 <td>596<td>10<td>BE419335</td></td></td>	596 <td>10<td>BE419335</td></td>	10 <td>BE419335</td>	BE419335
12 <td>56.6<td>2.2<td>611<td>9<td>AL506359</td></td></td></td></td>	56.6 <td>2.2<td>611<td>9<td>AL506359</td></td></td></td>	2.2 <td>611<td>9<td>AL506359</td></td></td>	611 <td>9<td>AL506359</td></td>	9 <td>AL506359</td>	AL506359
13 <td>56.6<td>2.2<td>487<td>10<td>BE490332</td></td></td></td></td>	56.6 <td>2.2<td>487<td>10<td>BE490332</td></td></td></td>	2.2 <td>487<td>10<td>BE490332</td></td></td>	487 <td>10<td>BE490332</td></td>	10 <td>BE490332</td>	BE490332
14 <td>56.6<td>2.2<td>544<td>14<td>BO294438</td></td></td></td></td>	56.6 <td>2.2<td>544<td>14<td>BO294438</td></td></td></td>	2.2 <td>544<td>14<td>BO294438</td></td></td>	544 <td>14<td>BO294438</td></td>	14 <td>BO294438</td>	BO294438
15 <td>55.8<td>2.2<td>632<td>13<td>BI895268</td></td></td></td></td>	55.8 <td>2.2<td>632<td>13<td>BI895268</td></td></td></td>	2.2 <td>632<td>13<td>BI895268</td></td></td>	632 <td>13<td>BI895268</td></td>	13 <td>BI895268</td>	BI895268
16 <td>55.4<td>2.2<td>461<td>12<td>BG263144</td></td></td></td></td>	55.4 <td>2.2<td>461<td>12<td>BG263144</td></td></td></td>	2.2 <td>461<td>12<td>BG263144</td></td></td>	461 <td>12<td>BG263144</td></td>	12 <td>BG263144</td>	BG263144
17 <td>55.4<td>2.2<td>517<td>13<td>BM137041</td></td></td></td></td>	55.4 <td>2.2<td>517<td>13<td>BM137041</td></td></td></td>	2.2 <td>517<td>13<td>BM137041</td></td></td>	517 <td>13<td>BM137041</td></td>	13 <td>BM137041</td>	BM137041
18 <td>55.4<td>2.2<td>806<td>10<td>BE040623</td></td></td></td></td>	55.4 <td>2.2<td>806<td>10<td>BE040623</td></td></td></td>	2.2 <td>806<td>10<td>BE040623</td></td></td>	806 <td>10<td>BE040623</td></td>	10 <td>BE040623</td>	BE040623
19 <td>55.2<td>2.2<td>444<td>14<td>BO490255</td></td></td></td></td>	55.2 <td>2.2<td>444<td>14<td>BO490255</td></td></td></td>	2.2 <td>444<td>14<td>BO490255</td></td></td>	444 <td>14<td>BO490255</td></td>	14 <td>BO490255</td>	BO490255
20 <td>55.2<td>2.2<td>466<td>13<td>BM099144</td></td></td></td></td>	55.2 <td>2.2<td>466<td>13<td>BM099144</td></td></td></td>	2.2 <td>466<td>13<td>BM099144</td></td></td>	466 <td>13<td>BM099144</td></td>	13 <td>BM099144</td>	BM099144
21 <td>54.8<td>2.2<td>823<td>14<td>BO804501</td></td></td></td></td>	54.8 <td>2.2<td>823<td>14<td>BO804501</td></td></td></td>	2.2 <td>823<td>14<td>BO804501</td></td></td>	823 <td>14<td>BO804501</td></td>	14 <td>BO804501</td>	BO804501
22 <td>54.6<td>2.2<td>469<td>14<td>BO470669</td></td></td></td></td>	54.6 <td>2.2<td>469<td>14<td>BO470669</td></td></td></td>	2.2 <td>469<td>14<td>BO470669</td></td></td>	469 <td>14<td>BO470669</td></td>	14 <td>BO470669</td>	BO470669
23 <td>54.4<td>2.2<td>681<td>14<td>BO768637</td></td></td></td></td>	54.4 <td>2.2<td>681<td>14<td>BO768637</td></td></td></td>	2.2 <td>681<td>14<td>BO768637</td></td></td>	681 <td>14<td>BO768637</td></td>	14 <td>BO768637</td>	BO768637
24 <td>54.4<td>2.2<td>700<td>9<td>AL504911</td></td></td></td></td>	54.4 <td>2.2<td>700<td>9<td>AL504911</td></td></td></td>	2.2 <td>700<td>9<td>AL504911</td></td></td>	700 <td>9<td>AL504911</td></td>	9 <td>AL504911</td>	AL504911
25 <td>54.2<td>2.2<td>776<td>12<td>BF624463</td></td></td></td></td>	54.2 <td>2.2<td>776<td>12<td>BF624463</td></td></td></td>	2.2 <td>776<td>12<td>BF624463</td></td></td>	776 <td>12<td>BF624463</td></td>	12 <td>BF624463</td>	BF624463
26 <td>54<td>2.1<td>478<td>13<td>BM442521</td></td></td></td></td>	54 <td>2.1<td>478<td>13<td>BM442521</td></td></td></td>	2.1 <td>478<td>13<td>BM442521</td></td></td>	478 <td>13<td>BM442521</td></td>	13 <td>BM442521</td>	BM442521
27 <td>54<td>2.1<td>638<td>14<td>BO293808</td></td></td></td></td>	54 <td>2.1<td>638<td>14<td>BO293808</td></td></td></td>	2.1 <td>638<td>14<td>BO293808</td></td></td>	638 <td>14<td>BO293808</td></td>	14 <td>BO293808</td>	BO293808
28 <td>53.8<td>2.1<td>477<td>9<td>AU164003</td></td></td></td></td>	53.8 <td>2.1<td>477<td>9<td>AU164003</td></td></td></td>	2.1 <td>477<td>9<td>AU164003</td></td></td>	477 <td>9<td>AU164003</td></td>	9 <td>AU164003</td>	AU164003
29 <td>53.6<td>2.1<td>516<td>13<td>BM078197</td></td></td></td></td>	53.6 <td>2.1<td>516<td>13<td>BM078197</td></td></td></td>	2.1 <td>516<td>13<td>BM078197</td></td></td>	516 <td>13<td>BM078197</td></td>	13 <td>BM078197</td>	BM078197
30 <td>53.6<td>2.1<td>563<td>10<td>BE498472</td></td></td></td></td>	53.6 <td>2.1<td>563<td>10<td>BE498472</td></td></td></td>	2.1 <td>563<td>10<td>BE498472</td></td></td>	563 <td>10<td>BE498472</td></td>	10 <td>BE498472</td>	BE498472
31 <td>53.6<td>2.1<td>852<td>12<td>BG310300</td></td></td></td></td>	53.6 <td>2.1<td>852<td>12<td>BG310300</td></td></td></td>	2.1 <td>852<td>12<td>BG310300</td></td></td>	852 <td>12<td>BG310300</td></td>	12 <td>BG310300</td>	BG310300
32 <td>53.4<td>2.1<td>935<td>17<td>CNS006XK</td></td></td></td></td>	53.4 <td>2.1<td>935<td>17<td>CNS006XK</td></td></td></td>	2.1 <td>935<td>17<td>CNS006XK</td></td></td>	935 <td>17<td>CNS006XK</td></td>	17 <td>CNS006XK</td>	CNS006XK
33 <td>53.2<td>2.1<td>338<td>13<td>BI306090</td></td></td></td></td>	53.2 <td>2.1<td>338<td>13<td>BI306090</td></td></td></td>	2.1 <td>338<td>13<td>BI306090</td></td></td>	338 <td>13<td>BI306090</td></td>	13 <td>BI306090</td>	BI306090
34 <td>53.2<td>2.1<td>576<td>14<td>BO458667</td></td></td></td></td>	53.2 <td>2.1<td>576<td>14<td>BO458667</td></td></td></td>	2.1 <td>576<td>14<td>BO458667</td></td></td>	576 <td>14<td>BO458667</td></td>	14 <td>BO458667</td>	BO458667
35 <td>53.2<td>2.1<td>578<td>14<td>BO458711</td></td></td></td></td>	53.2 <td>2.1<td>578<td>14<td>BO458711</td></td></td></td>	2.1 <td>578<td>14<td>BO458711</td></td></td>	578 <td>14<td>BO458711</td></td>	14 <td>BO458711</td>	BO458711
36 <td>53<td>2.1<td>562<td>12<td>BG263740</td></td></td></td></td>	53 <td>2.1<td>562<td>12<td>BG263740</td></td></td></td>	2.1 <td>562<td>12<td>BG263740</td></td></td>	562 <td>12<td>BG263740</td></td>	12 <td>BG263740</td>	BG263740
37 <td>53<td>2.1<td>935<td>17<td>CNS006XK</td></td></td></td></td>	53 <td>2.1<td>935<td>17<td>CNS006XK</td></td></td></td>	2.1 <td>935<td>17<td>CNS006XK</td></td></td>	935 <td>17<td>CNS006XK</td></td>	17 <td>CNS006XK</td>	CNS006XK
38 <td>52.8<td>2.1<td>472<td>12<td>BF585648</td></td></td></td></td>	52.8 <td>2.1<td>472<td>12<td>BF585648</td></td></td></td>	2.1 <td>472<td>12<td>BF585648</td></td></td>	472 <td>12<td>BF585648</td></td>	12 <td>BF585648</td>	BF585648
39 <td>52.8<td>2.1<td>522<td>9<td>AL505956</td></td></td></td></td>	52.8 <td>2.1<td>522<td>9<td>AL505956</td></td></td></td>	2.1 <td>522<td>9<td>AL505956</td></td></td>	522 <td>9<td>AL505956</td></td>	9 <td>AL505956</td>	AL505956
40 <td>52.8<td>2.1<td>552<td>10<td>AV940913</td></td></td></td></td>	52.8 <td>2.1<td>552<td>10<td>AV940913</td></td></td></td>	2.1 <td>552<td>10<td>AV940913</td></td></td>	552 <td>10<td>AV940913</td></td>	10 <td>AV940913</td>	AV940913
41 <td>52.8<td>2.1<td>563<td>13<td>BU479941</td></td></td></td></td>	52.8 <td>2.1<td>563<td>13<td>BU479941</td></td></td></td>	2.1 <td>563<td>13<td>BU479941</td></td></td>	563 <td>13<td>BU479941</td></td>	13 <td>BU479941</td>	BU479941
42 <td>52.8<td>2.1<td>569<td>14<td>BO485251</td></td></td></td></td>	52.8 <td>2.1<td>569<td>14<td>BO485251</td></td></td></td>	2.1 <td>569<td>14<td>BO485251</td></td></td>	569 <td>14<td>BO485251</td></td>	14 <td>BO485251</td>	BO485251
43 <td>52.8<td>2.1<td>593<td>13<td>BM317529</td></td></td></td></td>	52.8 <td>2.1<td>593<td>13<td>BM317529</td></td></td></td>	2.1 <td>593<td>13<td>BM317529</td></td></td>	593 <td>13<td>BM317529</td></td>	13 <td>BM317529</td>	BM317529
44 <td>52.8<td>2.1<td>594<td>10<td>AV934154</td></td></td></td></td>	52.8 <td>2.1<td>594<td>10<td>AV934154</td></td></td></td>	2.1 <td>594<td>10<td>AV934154</td></td></td>	594 <td>10<td>AV934154</td></td>	10 <td>AV934154</td>	AV934154
45 <td>52.8<td>2.1<td>616<td>14<td>BO789433</td></td></td></td></td>	52.8 <td>2.1<td>616<td>14<td>BO789433</td></td></td></td>	2.1 <td>616<td>14<td>BO789433</td></td></td>	616 <td>14<td>BO789433</td></td>	14 <td>BO789433</td>	BO789433

ALIGNMENTS

RESULT 1

LOCUS

DEFINITION

NCSM3H7T3 Subtracted Mycelial Neurospora crassa cDNA clone SM3H7 5' similar to avicelase III, Aspergillus aculeatus, mRNA sequence.

ACCESSION

BF072664.1 GI:10866169

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

JOURNAL

MEDLINE

COMMENT

Expressed sequences from conidial, mycelial, and sexual stages of Neurospora crassa
Fungal Genet. Biol. 21, 348-363 (1997)
Contact: Natvig, D.O./Nelson, M.A.
Department of Biology
University of New Mexico
Casterler Hall, Albuquerque, NM 87131, USA
Tel: 505 277 3411
Fax: 505 277 0304

BQ751758 426 bp mRNA linear EST 18-JUL-2002
LOCUS EST632321 DSCCT Colletotrichum trifolii cDNA clone pDSC7-37, mRNA
DEFINITION
ACCESSION BQ751758
VERSION BQ751758.1 GI:21907163
KEYWORDS EST
SOURCE Colletotrichum trifolii.
ORGANISM Colletotrichum trifolii.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes Incertae sedis; Phyllachorales; Phyllachoraceae;
mitosporic Phyllachoraceae; Colletotrichum.
REFERENCE
AUTHORS Samad, D.A., Dickman, M., Town, C.D., Van Aken, S., Uterback, T.,
Cheung, F., and Fraser, C.M.
TITLE ESTs from mycelia of Colletotrichum trifolii race 1
JOURNAL Unpublished (2002)
COMMENT Other-ESTs: EST632322
Contact: Deborah A. Samad
Department of Plant Pathology
University of Minnesota
495 Borlang Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
Tel: 612 625 1243
Fax: 651 649 5058
Email: debby@puccini.crl.umn.edu
TIGR sequence name: MTSAG37TK More information is available at:
www.medicago.org
Seq primer: SKmod (CTA gaa cta gtc gat cc).
Location/Qualifiers
1..426
/organism="Colletotrichum trifolii"
/strain="race 1"
/db_xref="taxon:5466"
/clone="pDSC7-37"
/clone_lib="pDSC7"
/russue_type="mycelia"
/dev_stage="young, actively growing mycelia (3 days after
inoculation) grown in liquid culture (cutin minimal medium
containing 2%glucose)."
/lab_host="DH5alpha"
/note="Vector: pBluescript SK+, site_1: EcoRI; site_2:
EcoRI; isolate: 2sp2; cDNA was prepared from polyA+
enriched RNA The cDNA was ligated into Lambda gtl1 from
Stratagene and packaged using Gigapack packaging extracts.
An aliquot of the amplified library was used to transduce
E. coli Y1090 and phage DNA was purified from a liquid
lysate. The cDNA inserts were gel purified after EcoRI
digestion and ligated into pBluescript SK+. Aliquots of
the ligation were used to transform E. coli DH5alpha which
were plated onto medium with X-gal for selection of
recombinants."

BASE COUNT 77 a 152 c 106 g 91 t
ORIGIN
Query Match 4.4%, Score 111.4; DB 14; Length 426;
Best Local Similarity 70.8%; Pred. No. 1.6e-14;
Matches 148; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 13 CGAGTCCTGCCCTGCTGCGGGCGCCGTCATCCCTGCCCATGCTGCTTTTATGAGAG 72
DB 186 CGAGGGTCGGGCTCTCTGCGGGCGCCGCTTCCTGTCGCGCGCTGCGACGTGGA 245
OY 73 AAGCTCAAGTCGCGGGCGGGCGGCTTCGTCGCCGATCATCTTCATCCCAAGACA 132
DB 246 AAGCTGACACAGCGGGGAGGCGCGGCTTCGTCGCGCTCATGCTTTCCACCCGACG 305
OY 133 AAAGCGTAGCATATGACGAAGATATTGGCGGGCTGTACCGGCTCAACGCCAGCAG 192
DB 306 AAGGCGCTGCGCTTCGCCCGAAGATATGCGGGCGCTGTACCGGCTCAACGCCAGC 365
OY 193 TCATGACCGCGCTTCAGCATGAGATTCG 221
DB 366 TCCTGTGACTCCGTCACGACGCAACGC 394

RESULT 4
CNS0091P/c 925 bp DNA linear GSS 03-JUN-1999
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053013.1 GI:4934461
VERSION AL053013
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS 1 (bases 1 to 925)
TITLE Genoscope.
JOURNAL Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org/TheBDGP/Drosophila>
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamonos in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1..925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPCI-98"
/note="end : TET3"

BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN
Query Match 2.6%, Score 66.4; DB 17; Length 925;
Best Local Similarity 15.4%; Pred. No. 0.00025;
Matches 56; Conservative 177; Mismatches 129; Indels 2; Gaps 1;

OY 2051 CTTGGGACGCGCGCGGCTGAGGGCTCGCTTACGCCAGTGAGACAGCGGCGCTCT 2110
DB 925 CSBSCSCSCSBSBSCSSSMSTSSNSBSCSSSSSTSSBSSBSSBSSBSSBSS 866
OY 2111 GGACGACATCCAGGAGCTCCAGGCGCTTCGCTCATGACAGCAACGAAGCTCCGCA 2170
DB 865 SSGTSSACVKNCASSCCCGCGMABCMCSSSSCCGSASARVYKVASGAGKRGGS 806
OY 2171 GCGGACAGACCGCGGGGAGCTAGTGGGACACACAGCGCGGCGCTTTTACGCTC 2230
DB 805 GGASASHSSSSACBSSSSSSCSASMSASSSSSRSGAGAGSASSSSSSSS 746
OY 2231 AGGAACCGTGGCGGCGGACAGGCGGAGCTCTCGTCGACCAAGACAGACACACA 2290
DB 745 ASAGSVSSASSSSSSSVSSSVSSASMSBSSSSASASSSSSSASACASCC 686
OY 2291 GTACCTTCCGCGACCTCGACACACAGCTGAGCTCAGCTTATTCACAGACCCGG 2350
DB 685 TSMSCGCTSMASMAARSSSSSSSSSSMSAS--SSASSSASASSSSSSSGSAG 628
OY 2351 CTTGACGCTGACTTCTGTCGAGACACACTCGCGCGCGGCTCCACGGGCTCAGGGCTC 2410

ACCESSION	B1305544
VERSION	B1305544.1
KEYWORDS	GI:14980866
SOURCE	EST.
ORGANISM	Oryza sativa.
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartioideae; Oryzaceae; Oryza.
TITLE	1 (bases 1 to 458)
JOURNAL	Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabushan,I., Ravindrababu,P. and Bennetzen,J.L. Novel EST enrichment with normalized cDNA libraries from drought stressed rice (Oryza sativa L.cv Nagina 22) Unpublished (2001)
COMMENT	Contact: Reddy AR Department of Plant Sciences, School of Life Sciences University of Hyderabad P.O. Central University, Hyderabad-500 046, A.P., India Tel.: 0091-40-3010265 Fax: 0091-40-3010145 Email: arjulsleuohyd.ernet.in Insert length: 458 Std Error: 0.00 Plate: 0 row: 1 column: 03 Seq primer: GTAAACGACGCCCATC. Location/Qualifiers .458 /organism="Oryza sativa" /cultivar="Nagina 22 [Indica sub sp]" /db_xref="taxon:4530" /clone="NL_0.103" /clone_lib="Drought stress (leaf)" /tissue_type="Entire leaf tissue" /dev_stage="35 day-old seedlings" /note="Organ: Leaf; Vector: T773pac; ESTs from normalized leaf cDNA library from drought stressed seedlings"
BASE COUNT	89 a 176 c 138 g 55 t
ORIGIN	
Query Match	2.5%; Score 62.4; DB 13; Length 458;
Best Local Similarity	53.9%; Pred. No. 0.0015;
Matches 151; Conservative 0; Mismatches 126; Indels 3; Gaps 1;	
OY	1523 AGAGGTGGTCCGCGGCACACACC GCCGACGCCAAGAGTG GCCCATCTGTC CGAGC 1582
DB	458 ACACGGCGACGCGCGCGGACGCGCATGCGCGCGCTGAGACGTGATGCGTTCCG 399
OY	1583 GCGGCGCACGTGAGACATCACTACGCGGCGGACACGTCATTGAACGGCGACAGTTG 1642
DB	398 CCGCGCGCATGTCGTTCCAGACTGTCATGTGTAATGCATCTGTCAGACACTGGCTCG 339
OY	1643 CCTATTGGCGGACGCGCGACGATCTCTTGTCGACCGCGCTGTCGCGGCGTGAACGCT 1702
DB	338 CCGGCGTGGTGCGCCCGCACACAG---CTGGTGCAGGCGCTGTCGTCGCGCGCGAC 282
OY	1703 GCGAGTTCACGGGACGCTTGGCTCGGTCGAGCTGCCCGCGGCGCGGCATCGCCT 1762
DB	281 TGCTGCTTTTCGACAGCGCGCGACGCGCATGTCGTCGCGCGTGTGAGCGAG 222
OY	1763 CGGACAGAAGACAACAGCGCTTCTAAGCGCGGCTCCG 1802
DB	221 AGGACAGAAGAACGCGGAGGAGTCTGTCGCGCGCGGAGATG 182
RESULT 8	
CNS0091P	
LOCUS	CNS0091P 925 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TRF3 end of BAC # BACR1916 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL053013
VERSION	AL053013.1
KEYWORDS	GI:4934461

REFERENCE	TITLE	AUTHORS	JOURNAL
SOURCE	Drosophila melanogaster.		
ORGANISM	Drosophila melanogaster.		
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
	Ephyridioidae; Drosophilidae; Drosophila.		
	1 (bases 1 to 925)		
COMMENT	<p>Genoscope.</p> <p>Direct Submission</p> <p>Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr</p> <p>- Web : www.genoscope.cns.fr)</p> <p>Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.</p> <p>Location/Qualifiers</p> <p>1..925</p> <p>/organism="Drosophila melanogaster"</p> <p>/db_xref="taxon:7227"</p> <p>/clone="BACR19D16"</p> <p>/clone_11b="RPc1-98"</p> <p>/note="end : TET3"</p>		
FEATURES	<p>source</p>		
BASE COUNT	120 a	61 c	61 g
ORIGIN	172 t 511 others		
Query Match	2.4%	Score 60.8:	DB 17; Length 925:
Best Local Similarity	14.0%	Pred. No. 0.0045;	
Matches	52; Conservative	173; Mismatches	147; Indels 0; Gaps 0;
Db	1504	TACGCCGGGAACCTCGGTCAAGAGCGTCGTCGCCGCGGACACACCGCGGACGAACAG	1563
Db	553	TTSSSGGKGGKSSSGSGBSCSCSSCSGCCSCCSCCSCCSCCSCGSSGSSSSSSKSS	612
Qy	1564	GTGCCATCTCGTCCGACGCGCGCGCGACGTGACATGACTACGCGGCGACACGTCC	1623
Db	613	TSBSCGCCSCSKSVCGTSCSSSSSSSSSTSSSTSSSTSSSKSSSGSSSSSSSYTTS	672
Qy	1624	ATGAACGGGGGACGCGTGGCCATTTCGGCGGACGGGACACATCCTCGTGCACCGCC	1683
Db	673	KSTASAGSSSWAGGSGSTGTTSSSSSSSTSSSSSVSGSSKSTBSBSSGSSSSSS	732
Qy	1684	TGCTCCGGGTCACACGTCGCGATGTCACAGGCGACGCTTTCCTCGCTCGACCTGCCC	1743
Db	733	SSSTSBBSCTSTSSSSSSSYSSSTSCCTCCCSYSSSTSSSTSSSTSMGSSSSSV	792
Qy	1744	CGCGGCGCGCTATCGCTCGGACGAAGAACAACAGCGTCTTACGCGCGGCTCCGA	1803
Db	793	GTSSSSDSTSTCCSCCYMCCTCTYBMCYNTSCGGSSSSGKGGYKCCCGCGSSS	852
Qy	1804	TGACCTTTTACGTAGCAGAGACACCGGACACAGTTTACGCGGCGGCGCAAGCTGGCC	1863
Db	853	TGCMGTTSSACSSSSSCSSSSSVSSSSSKSSAASSSSSVSSGSSGVSNSSSASKSSSSG	912
Qy	1864	ACGCGACGACG 1875	
Db	913	SSGSGSGSGSV 924	
RESULT 9	BQ463043 559 bp mRNA linear EST 30-MAY-2002		
LOCUS	BQ463043		
DEFINITION	H102N16r HI Hordeum vulgare cDNA clone H102N16 5-PRIME, mRNA		

ACCESSION	sequence.
VERSION	BQ463043
KEYWORDS	BQ463043.1 GI:21270825
SOURCE	EST.
ORGANISM	<i>Hordeum vulgare</i> .
	<i>Hordeum vulgare</i> .

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (pages 1 to 559)	Zhang, H., Wesschke, W., Michalek, W., Stein, N. and Graner, A	EST sequencing and analysis in barley (2002)	Unpublished (2002)	Contact: Stein Nils

Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel.: 039482-5522
Fax: 039482-5595
Email: scotainipk-gatersleben.de
Insert length: 559 Std Error: 0.00
plate: 2 row: N column: 16
Seq primer: M3rev.

FEATURES	location/Qualifiers
source	1. .559
	/organism="Hordoun"

```

/organism="Homo sapiens"
/guiltivar="Barke"
/db_xref="taxon:4513"
/clone="H102N16"
/clone_1lb="HI"
/tissue.type="female inflorescences"
/dev_strage="female inflorescences (approx. 3 mm in size)"
/lab_host="X110-Gold"
/note="Vector: pBluescript SK+, Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning artifact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream of EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used, Blue/white selection for recombinants is not 100% reliable."

```

BASE COUNT	75 A	279 C	122 G	83 T
ORIGIN				
Query Match		2.48;	Score 60.2;	DB 14;
Best Local Similarity		46.5%;	Pred. No. 0.0051;	Length 559;
Matches 194; Conservative		0;	Mismatches 223;	Indels .0;
				Gaps 0;

[illegible]

db 200 ACAACCGCGCTCTCCCTCTGCTTGCACCTCCGTCACACCCACCGCGCTTCCTCCGA 259

Db 260 CCGTCCCTCCCTCTCTTCGACCAACCCTACCGGCCACCAAGCTCCAGTTCACACCGC 319
07 1592 CGGAGGACATTCGACATACGCGGCGGACACAGTCCATCAAGACGGCGGCGGTGGCTATTTCGG 1651

1652 CCGACGGCGACAGATCTTGGTGGACCGCCCTGCTCCGGCGTGCAGCGGTCGACATTCC 1711

Db 320 GCGCGCCCTCCACGCCCCCTCCTCGCCCTCTCCCTCGAGCGGCCCTCGGCTTGGCAGCCCC 379

oy 1712 AGGCGAGCTTTGCCCTCCGTCTGAGCCTGCACGGGGGGCGCGCATTCGCCCTGGACAAGA 1771

nb 380 CCCCTGGAGCAACTTCTCCGCTTCCCCCGGGGAGGCTCCGCTTGCTCTCTGGACAACC 439

22 1772 AGACCAACAGCGCTCTTCTACGGCCGGGCTCCGGATGCGACCTTTTACGTGACGACGAAGACACCG 1831

Db 440 GCAAGGCTCCGCGCTGGAGTTCCTGCGCGCCCTCACCCTCTCGACTGGAACGAATCG 499

1832 GCAGCAGCTTTCACGCGCGGGGCCCAAGCTGGGCAAGCGCAGGAGCAGCATCCGGGATATCG 1888

Db 500 AGCCCGCGCATCGGACCGCTCCATCGACACACCTGCACCGTCTGGACATCG 556

LOCUS	629 bp	mrna	linear	EST 28-JUN-2002
DEFINITION	Tar1134H10F Tar11 triticum aestivum cDNA clone Tar1134H10F, mRNA sequence.			

ACCESSION	BO620053
VERSION	BO620053.1
KEYWORDS	GI:21625132
SOURCE	EST.
ORGANISM	bread wheat. <i>Triticum aestivum</i>

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 629)
Cloutier,S., Dong,G., and Walsh,A.
Wheat functional genomics - Thatcher Lr1 cDNA library
Unpublished (2001)
Contact: Dr. Sylvie Cloutier
Ouellet-Bacot, Genevieve; Bouchard, and Marie-Fred, Canada

Eupariota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta,
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Triticum.

CONTACTS

Cereal Research Centre, Agriculture and Agri-Food Canada
195 Dafoe Rd., Winnipeg, MB, Canada R3T 2M9
Tel.: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca

was cloned directionally, not all sequences generated with reverse primer where from the 5' end (same with forward primer and 3'end).
Average insert size is >2.kbp
Plate: 134 row: H column: 10
Seq primer: M13 Forward.
location/qualifiers

```

SOURCE
LOCATIONS/Vectors
1. 629
/organism="Triticum aestivum"
/cultivar="Hatcher Lr1"
/db_xref="taxon:4565"
/clone="TaLr134H10F"
/clone_lib="TaLr1"
/tissue_type="leaf tissue"
/dev_stage="14 days old"
/lab_host="E. coli K10LR"
/note="Vector: Lambda ZapII; mass excised in plasmid

```

BASE COUNT	88 a	233 c	196 g	112 t
ORIGIN	vector pBK-CMV (Stratagene).; Site_1: EcoRI; Site_2: XhoI; mRNA obtained from wheat NIL Thatcher Lrl 24 hours after inoculation with leaf rust pathogen Puccinia triticina race BB8 carrying the avirulence gene Avr1."			

Query Match	2.3%	Score 59	DB 14	Length 629
Best Local Similarity	45.3%	Pred. No. 0.0099		
Matches 215; Conservative	0	Mismatches 260	Indels 0	Gaps 0
QY 1748	GGCGGTCATGCGCCGCGACAGAGACCAACAGCGCTTTCACGCGGCGCTCGGATCGA	1807		

Db 131 GCGCCGCGCTCGCCGAGTTCACTTCATGCTCACTCTCGTCTTGGCCGCGCTCAGGCTCCG 190

QY 1808 CCTTTACGTCAGCAGGACACCGGACAGCTTCACCGCGGGGCCAACCTGGCCAGCG 1867
 Db 191 GCATGGCCCTTGGCGACGCTGATGACGGGGTGGACGACGACGCGGGGGGGGGTGGATGGCG 250

20 CAGGACGATCCGGGATATCCGTGCTCACCCGACACCCCGGGCAGTTTGATGTTCTCGA 1927

Db 251 C G G C C T G G C C C A C G C G T T C G G C C T T T C G T G G C C G T C T C C G T G G G G C C C A C A T C T C G G 310

Qy 1928 CCGAGCTGGCATATTCGCTCCACAGACTGGGACGACCTTTGGCCAAGTCTCCACCG 1987

DB 311 GCGGTCACGTAACCCGGGCGCGTACATTCGCGCGCTTCGTAAGGCGGACACATCAAGCCCTC 370

QY 1988 CCCTGCACCAACACCTTACAGATCGCCCTGGGTGTGGGCTCAGGCTCGAAGCTGAACCTGT 2047

Db 371 TCAAGCGGTGCTGTTACTGGGTGGCGCAGCTGCTCGGCTCCGTGCTGGCCCTGCTCTCC 430

Oy	2048	ATGCTTGCGGACGGCCCGGTCAAGGGGGTTCGGCTTCAGGCCAGGTGAGACAGCGGGCGCT	2107
Dd	431	TCAAGATGCGCACCGGCGGCGAGCGCCGTGGGCGCTTTCTTCGCTTCGCCGCGCGTGGGGC	490
Oy	2108	CCTGAGCGAGATCCAGAGGCGTCCCAAGGGCGTTGCCTCATCGACAGACAAGTAGTGGCG	2167
Dd	491	TCTGGAAAGCGCGTGCTTTCGAGATCGATCATCACTTCAGGCGCTCGTTCACACCGTCTAAG	550
Oy	2168	GCAGCGGACAGACCGCGGCGCAAGTCTACGTAGGACACACAGCGCGCGGCGCTT	2222
Dd	551	CCACGCGCGTGCACCCCAAGCGCGGCGAGACTCGGTATCGCGCGCCATCGCCAT	605
 RESULT 11 BEA19335 LOCUS BEA19335 DEFINITION MMR06.G9R000101 ITFC WMR Wheat Root Library Trilicium aestivum cDNA ACCSSION BEA19335 VERSION BEA19335 KEYWORDS BEA19335.1 GI:9417181 SOURCE EST. ORGANISM bread wheat. Trilicium aestivum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triliceae; Trilicum. 1 (bases 1 to 596) REFERENCE Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier, S., Dubcovsky,O., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemijn,J.M., Jia,J., Joulier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y., Pechion,L.N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Weizel,G.: Production of International Triliceae EST cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triliceae Unpublished (2000) JOURNAL Contact: Schuch W COMMENT Zeneca Wheat Improvement Centre, Norwich Research Park Colney Lane, Norwich NR4 7UH UNITED KINGDOM Tel: 44 1603 250 2600 Fax: 44 1603 250 699 Email: wolfgang.schuch@wageningen-zeneca.com International Triliceae EST Cooperative (ITEC) http://wheat.pw.usda.gov/genome. location/qualifiers			
FEATURES Source 1..596 /organism="Trilicium aestivum" /cultivar="Novosibirskaya 67" /db_xref="taxon:4565" /clone="WMR06.G9" /clone_lib="ITEC WMR Wheat Root Library" /tissue_type="root" /note="M13 Reverse sequencing primer used for 5' end of clone."			
BASE COUNT 88 a 271 c 172 g . 65 t			
ORIGIN			
Query Match 2.2% Score 56.6; DB 10; Length 596; Best Local Similarity 44.9%; Pred. No. 0.033; Matches 258; Conservative 0; Mismatch 314; Indels 3; Gaps 1;			
Oy	1588	GCGAGGTGAGCATGATGATGACGGCGGCGACAGTCAATGAACGCGGACGAGTGGCTTAT	1647
Dd	19	GCGAGGAGGAGGCGCGGCTACGAGGCGGACCAAGCGACGCGCGCGCCACAGTGGCGTC	78
Oy	1648	TGCGCCGAGCGGACACAGATCTCTGTGTGACGCGCTCGTCCGCGTCAAGCGCTCG---	1704
Dd	79	GCGCGCGCGCGGACCTCGCGCTCGCGCGCTTCGCGACGAGCGCGCTTCGCGCGCGG	138
Oy	1705	CAGTTCACAGGCGAGCTTTCGCTTCGCTTCGACGCTTCGCGCGCGCGCGCGCTATGCGCTCG	1764
Dd	139	CCCTCTTCGACGGGCGCTACCTCACTCACTCGAGAAGCCGCGCGCTTCTCTCGACCTC	198

Oy	1765	GAAAGACGACCAAGGGCTCTTTCACGCCCGGCTCCGGATGCACCTTTTACTCAGCAAG	1824
Db	199	AAGCCCCAACACAGATGTGTGCTTCAGATTTCATGAATCGGCGACGGTGCTCGACAAG	258
Oy	1825	GACACCGGACAGCTTATCAGCGCGCGGCCCAAAGCTGGGACGCGCACAGATTCGGGAT	1884
Db	259	CGGGTCAGGCTACCTACACGACACTTCACACTCCCTCGCGCCGCGCGCCAAGCCGCGC	318
Oy	1885	ATCGCTGCTACCCGACACACCGCGGGGCAAGCTTGATGTCTCGACCGAGCTCGCATATTC	1944
Db	319	ATCCCGGGCCCCGGCGCGCGCGCCGCGCCGCGCCCTTAGAGGGGCCCGCTACC GGCG	378
Oy	1945	CGCTCCACAGACCTCGGGGACGACCTTTTGGCCAAAGTCTCACCCGCCCTAGCAACACATAC	2004
Db	379	CGCAGCGGGCTCGACCTCTTCATCCCTTCGACCCCGCCACAAAGATCAGCTTCGCCAC	438
Oy	2005	CAGATCGCCCTGGTGTGTGGGCTCAGGCTTCGAAGTGAACCTTGATCCCTTCGACACGGC	2064
Db	439	GGCGTCGGCGGGGGGGGGCGCGGGTCAAGTACACTTAAGGGCACAGGGCGCGCGGCTC	498
Oy	2065	CCGTACGGGGCTCGCCTCTACGCCAGTGAAGACAGCGGGCGCTCTCGACGACATCCAG	2124
Db	499	ACCACCATTCGAGCCCTGTACGACACCGCCCAAGAAGCGCTGGAGATTGCGCTGCCACAG	558
Oy	2125	GGCTCCCAAGGGGCTTGCGCTCATATCACACGACACCA	2159
Db	559	AAGTTTCAGCCCGGGACACCTCTCAGGGGCACTTA	593
RESULT 12			
AL506359/c			
LOCUS	AL506359	611 bp	mRNA linear EST 04-JAN-2001
DEFINITION	AL506359 Hordeum vulgare Barke developing caryopsis (3..15.DAP)		
VERSION	AL506359		
KEYWORDS	AL506359.1 GI:12032574		
SOURCE	EST.		
ORGANISM	Hordeum vulgare.		
	Eukarya; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae		
	; Triticeae; Hordeum.		
	1 (bases 1 to 611)		
REFERENCE	Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.		
AUTHORS	EST sequencing and analysis in barley		
TITLE	unpublished (2000)		
JOURNAL	Contact: Michalek W		
COMMENT	Institute for Plant Genetics and Crop Plant Research Corrensstr.3, D-06466 Gatersleben, Germany Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de Seq primer: T3 primer for 5'end. Location/Qualifiers 1..611 /organism="Hordeum vulgare" /cultivar="Barke" /db_xref="taxon:4513" /clone="HY02N19T" /clone_lib="Hordeum vulgare Barke developing caryopsis (3..15.DAP)" /tissue_type="developing caryopsis (3..15.DAP)" /lab_host="XLOLR" /note="Vector: plasmid pBK-CMV. Site.1: EcoRI; Site.2: XhoI; mRNA was made from developing caryopsis (3..15 DAP) of spring barley variety 'barke', a high quality malting variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI (3'-end of cDNA). NOTE: Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter. Average insert size is 1 kb Sequence trimming: Vector sequences and sequence ends were trimmed from the 5'-and 3'-end until a 50 bp window contains less than two ambiguities. The maximum length was set to 700 bp"		

Wed May 7 14:15:40 2003

Search completed: May 4, 2003, 08:22:51
Job time : 2325.53 secs

us-10-026-994-4.rst

Page 10